



⑯ BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

⑯ Übersetzung der  
europäischen Patentschrift

⑯ EP 0 539 358 B1

⑯ DE 690 30 959 T 2

⑯ Int. Cl. 6:  
G 11 C 11/56  
G 11 C 16/06

DE 690 30 959 T 2

⑯ Deutsches Aktenzeichen: 690 30 959.7  
 ⑯ PCT-Aktenzeichen: PCT/US90/01984  
 ⑯ Europäisches Aktenzeichen: 90 906 720.9  
 ⑯ PCT-Veröffentlichungs-Nr.: WO 90/12400  
 ⑯ PCT-Anmeldetag: 12. 4. 90  
 ⑯ Veröffentlichungstag der PCT-Anmeldung: 18. 10. 90  
 ⑯ Erstveröffentlichung durch das EPA: 5. 5. 93  
 ⑯ Veröffentlichungstag der Patenterteilung beim EPA: 18. 6. 97  
 ⑯ Veröffentlichungstag im Patentblatt: 27. 11. 97

⑯ Unionspriorität:

337579 13.04.89 US  
 508273 11.04.90 US

⑯ Patentinhaber:

SunDisk Corp., Santa Clara, Calif., US

⑯ Vertreter:

Hoffmann, E., Dipl.-Ing., Pat.-Anw., 82166 Gräfelfing

⑯ Benannte Vertragstaaten:

DE, FR, GB, IT, NL

⑯ Erfinder:

MEHROTRA, Sanjay, Milpitas, CA 95035, US;  
 HARARI, Eliyahou, Los Gatos, CA 95035, US; LEE,  
 Winston, San Francisco, CA 94122, US

⑯ EEPROM mit Referenzzelle

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelebt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II 53 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patentamt inhaltlich nicht geprüft.

DE 690 30 959 T 2

EP 0 539 358

5

### Hintergrund der Erfindung

Diese Erfindung bezieht sich allgemein auf elektrisch löschbare programmierbare Halbleiter-Festwertspeicher (EEPROM) und betrifft insbesondere Schaltungen und Techniken zum Lesen und Programmieren des Zustandes derselben.

EEPROM sowie elektrisch programmierbare Festwertspeicher (EPROM) werden typischerweise in digitalen Schaltkreisen für nichtflüchtiges Speichern von Daten oder Programmen benutzt. Sie können gelöscht werden, und neue Daten werden in ihre Speicherzellen geschrieben oder "programmiert".

Ein EPROM arbeitet mit einem schwebenden bzw. schwimmenden (unverbundenen) leitenden Gate in einer Feldeffekttransistorkonstruktion, welches oberhalb einer Kanalzone in einem Halbleitersubstrat, aber isoliert gegenüber derselben zwischen Source- und Drainzonen angeordnet ist. Über dem schwimmenden Gate ist dann ein Steuergate vorgesehen, aber gleichfalls gegenüber diesem isoliert. Die Schwellenspannungskennlinie des Transistors wird durch die Ladungsmenge gesteuert, die am schwimmenden Gate gehalten wird. Das bedeutet, daß die Mindestspannungsmenge (Schwelle), die an das Steuergate angelegt werden muß, ehe der Transistor durchgeschaltet wird, um ein Leiten zwischen seiner Source- und Drainzone zu ermöglichen, vom Spannungspegel des schwimmenden Gates gesteuert ist.

Das schwimmende Gate kann einen Bereich an Ladung halten, und deshalb kann eine EPROM-Speicherzelle auf einen beliebigen Schwellenpegel innerhalb eines Schwellenfensters programmiert werden. Die durch den Mindest- und Höchstschwellenpegel der Vorrichtung bestimmte Größe des Schwellenfensters hängt von den Kennlinien der Vorrichtung, den Betriebsbedingungen und der Entstehung ab. Im Prinzip kann jeder einzelne Schwellenpegel innerhalb des Fensters zum Bezeichnen eines bestimmten Speicherzustandes der Zelle benutzt werden.

Zur Speicherung im EPROM wird der als Speicherzelle dienende Transistor dadurch auf einen von zwei Zuständen programmiert, daß Elektronen aus der Kanalzone im Substrat durch ein Düringatedielektrikum und auf das schwimmende Gate beschleunigt werden. Die Speicherzustände sind dadurch löschar, daß die Ladung vom schwimmenden Gate mittels Ultraviolettbestrahlung entfernt wird.

Ein elektrisch löscharer und programmierbarer Festwertspeicher (EEPROM) hat einen ähnlichen Aufbau, bietet aber zusätzlich einen Mechanismus zum Entfernen von Ladung von seinem schwimmenden Gate, wenn entsprechende Spannungen angelegt werden. Ein Feld solcher EEPROM-Zellen wird als "Flash"-EEPROM-Feld bezeichnet, wenn ein ganzes Feld von Zellen

oder eine bedeutende Gruppe von Zellen des Feldes gleichzeitig (d.h. blitzartig) gelöscht wird. Dann kann eine Zelle, sobald sie gelöscht ist, neuprogrammiert werden.

5 Eine bestimmte einzelne Zelle in einem zweidimensionalen Feld aus EPROM, EEPROM-Zellen wird zum Lesen dadurch adressiert, daß eine Source-Drainspannung an die Source- und Drainleitung in einer die zu adressierende Zelle enthaltenden Spalte angelegt wird, und daß eine Steuergatespannung an eine Wortleitung angelegt wird, die mit den Kontrollgates in einer die zu adressierende Zelle enthaltenden Reihe verbunden ist.

10 Der Zustand des Transistors einer adressierten Speicherzelle wird dadurch gelesen, daß eine Betriebsspannung bei ihm zwischen Source und Drain und an sein Steuergate angelegt und dann der Pegel des zwischen Source und Drain fließenden Stroms erfaßt wird. Der Strompegel ist proportional dem Schwellenpegel des Transistors, der seinerseits durch die Ladungsmenge an seinem schwimmenden Gate bestimmt ist.

15 In der üblichen EEPROM-Zelle mit zwei Zuständen wird ein Trennschwellenpegel festgelegt, damit das Schwellenfenster in zwei Bereiche unterteilt werden kann. Der Source/Drain-Strom wird mit dem Trennschwellenpegel verglichen, der beim Programmieren der Zelle benutzt wurde. Ist der abgelesene Strom größer als der der Schwelle wird bestimmt, daß sich die Zelle in einem "0"-Zustand befindet, während bei einem Strom, der geringer ist als der der Schwelle, die Zelle als in dem anderen Zustand befindlich festgelegt wird. Eine solche Zweizustandszelle speichert also ein Bit digitaler Information. Häufig wird als Teil eines Speichersystems eine Stromquelle vorgesehen, die von außen programmierbar sein kann, um den Trennschwellenstrom zu erzeugen.

20

25 IEEE JOURNAL OF SOLID-STATE CIRCUITS, Bd. SC-22, Nr. 4, Aug. 1987, New York, USA, Seiten 548-552, MASUOKA et al.: "A 256-kbit Flash EEPROM Using Triple-Polysilicon Technology" offenbart einen Flash-EEPROM gemäß dem Oberbegriff von Anspruch 1. Ein EEPROM mit einer Vielzahl verschiedener Schwellen zum Unterscheiden von vier Zuständen von Mehrfachbitzellen ist bekannt aus IEEE JOURNAL OF SOLID-STATE CIRCUITS, Bd. SC-22, Nr. 3, Juni 1987, New York, USA, Seiten 460-463, BLEIKER et al. "A Four-State EEPROM Using Floating-Gate Memory Cells", US-A-4 460 982, Gee et al. "Intelligent Electrically Programmable and Electrically Erasable ROM", US-A-4 495 602, Sheppard, D.P. "Multi-bit Read Only Memory Circuit".

30

35 Im Fall einer EEPROM-Speicherzelle mit Mehrfachzuständen speichert folglich jede Zelle zwei oder mehr Bits an Daten. Dadurch erweitert sich die Information, die ein gegebenes EEPROM-Feld speichern kann, um das Mehrfache der Anzahl von Zuständen, die jede Zelle speichern kann.

40 Dementsprechend ist es eine Hauptaufgabe der vorliegenden Erfindung, ein System aus EEPROM-Speicherzellen zu schaffen, welches es möglich macht, die Zellen zum Speichern von mehr als einem Bit Daten zu benutzen.

Eine weitere Aufgabe der vorliegenden Erfindung besteht in der Schaffung verbesserter Lese-  
schaltkreise als Teil eines EPROM- oder EEPROM-Speicherchips in integrierter Schaltung.

Eine Aufgabe der Erfindung besteht auch darin, Leseschaltkreise zu schaffen, die einfacher sind,  
5 sich leichter herstellen lassen und verbesserte Genauigkeit und Zuverlässigkeit über eine lange  
Benutzungsspanne haben.

Es ist auch eine Aufgabe der vorliegenden Erfindung, es möglich zu machen, verbesserte  
Programmschaltkreise als Teil eines EPROM- oder EEPROM-Speicherchips in integrierter Schal-  
10 tung bereitzustellen.

Eine Aufgabe der Erfindung ist es auch, es möglich zu machen, Programmierschaltkreise zur  
Verfügung zu stellen, die einfacher sind, sich leichter herstellen lassen, und verbesserte  
Genauigkeit und Zuverlässigkeit über einen langen Benutzungszeitraum haben.

15 Eine weitere Aufgabe der vorliegenden Erfindung besteht in der Bereitstellung von Speicherlese-  
und -programmiertechniken, die Auswirkungen der Temperatur, Spannungs- und Prozeßschwan-  
kungen sowie des Haltens von Ladung automatisch ausgleichen.

20 Noch eine weitere Aufgabe der vorliegenden Erfindung besteht in der Schaffung von Flash-  
EEPROM-Halbleiterchips, die Speichervorrichtungen in Form von Magnetplatten in Rechnersy-  
stemen ersetzen können.

Ferner ist es eine Aufgabe der vorliegenden Erfindung, eine Flash-EEPROM-Konstruktion zu  
25 schaffen, die zu einer erhöhten Lebenszeit, gemessen anhand der Zahl Programmier/Lesezyklen,  
die der Speicher aushalten kann, fähig ist.

Gemäß der Erfindung wird ein nichtflüchtiger Speicher geschaffen, umfassend: ein Feld einer  
Vielzahl adressierbarer, elektrisch löscherbarer und programmierbarer Halbleiter-Speicherzellen der  
30 Art, die eine Source, eine Drain, ein Steuergate, ein schwimmendes Gate, welches in der Lage  
ist, einen bei der Benutzung des Speichers programmierten Ladungsspegel zu halten, und eine  
Löschelektrode aufweist, welche in der Lage ist, Ladung von dem schwimmenden Gate zu  
entfernen. Gleichfalls eingeschlossen ist ein System zum Lesen, ob der programmierte Zustand  
35 einer adressierten Zelle größer oder kleiner ist als eine vorherbestimmte Schwelle, eine Refe-  
renzspeicherzelle sowie eine auf die Referenzzelle reagierende Leseeinrichtung zum Vergleichen  
des Ladungsspegels einer adressierten Zelle mit dem der Referenzzelle. Bedeutsamerweise gehört  
zu diesem Speicher eine auf Löschen und Neuprogrammieren des Speicherzellenfeldes reagie-  
rende Einrichtung zum Löschen und Neuprogrammieren der Referenzzelle mit einer Ladung, die  
im wesentlichen gleich oder proportional der Schwelle ist.

40 Ferner wird gemäß der Erfindung ein nichtflüchtiger Speicher geschaffen, umfassend ein  
EEPROM-Feld, welches eine Vielzahl von Gruppen einzeln adressierbarer EEPROM-Speicherzel-  
len enthält, wobei jede Zellengruppe zusammen als eine Einheit löscherbar ist, mit mindestens  
einem vorherbestimmten Schwellenpegel zur Abgrenzung zwischen zwei Leitzuständen der

Zellen in dem EEPROM-Feld, sowie mindestens eine Gruppenreferenzzelle, die als Teil jeder Speicherzellengruppe zum Speichern eines dem vorherbestimmten Schwellenpegel entsprechenden Gruppenschwellenpegels vorgesehen ist. Vorgesehen ist auch eine Einrichtung zum Lesen einer adressierten Zelle der Zellengruppe durch Vergleich mit dem Gruppenschwellenpegel, sowie eine auf das Löschen und Neuprogrammieren der Zellengruppe reagierende Einrichtung zum Löschen und Neuprogrammieren der Gruppenreferenzzelle auf einen dem vorherbestimmten Schwellenpegel entsprechenden Gruppenschwellenpegel.

10 **Kurzbeschreibung der Zeichnungen**

Fig. 1 ist eine Querschnittsansicht einer EEPROM-Vorrichtung als integrierte Schaltungskonstruktion, die sich zur Verwirklichung der verschiedenen Aspekte der vorliegenden Erfindung eignet;

15 Fig. 2 ist eine Ansicht der in Fig. 1 gezeigten Konstruktion längs des Schnitts 2-2;

Fig. 3 ist ein Ersatzschaltkreis einer einzigen EEPROM-Zelle des in Fig. 1 und 2 dargestellten Typs;

20 Fig. 4 zeigt ein adressierbares EEPROM-Zellenfeld;

Fig. 5 ist ein Blockschaltbild eines EEPROM-Systems, bei dem die verschiedenen Aspekte der vorliegenden Erfindung verwirklicht sind;

25 Fig. 6 veranschaulicht die Aufteilung des Schwellenfensters einer EEPROM-Zelle, die ein Bit Daten speichert;

30 Fig. 7A veranschaulicht die Aufteilung des Schwellenfensters einer EEPROM-Zelle, die zwei Bits Daten speichert;

Fig. 7B veranschaulicht die Aufteilung des Schwellenfensters des Leitungsstroms von Source zu Drain bei der EEPROM-Zelle gemäß Fig. 7A;

35 Fig. 8A und 8B sind Kurven, die die Änderungen und Kennlinien eines typischen EEPROM nach einer Nutzungsdauer darstellen;

Fig. 9A veranschaulicht Lese- und Programmierschaltungen für eine Master-Referenzzelle und eine adressierte Speicherzelle gemäß der vorliegenden Erfindung;

40 Fig. 9B veranschaulicht Leseschaltkreise mit Mehrfachzuständen und mit Referenzzellen gemäß der vorliegenden Erfindung;

Fig. 9C(1)-9C(8) veranschaulichen die Zeitbestimmung für das Mehrfachzustandslesen für die in Fig. 9B gezeigten Schaltkreise;

5 Fig. 10 stellt eine spezielle Speicherorganisation gemäß einem Ausführungsbeispiel der vorliegenden Erfindung dar;

Fig. 11 zeigt einen Algorithmus zum Programmieren eines Satzes örtlicher Referenzzellen gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

10 Fig. 12A zeigt ein Ausführungsbeispiel eines Leseschaltkreises, der örtliche Referenzzellen direkt benutzt;

Fig. 12B zeigt einen Lesealgorithmus für das Ausführungsbeispiel gemäß Fig. 12A;

15 Fig. 13A zeigt ein alternatives Ausführungsbeispiel eines Leseschaltkreises, der örtliche Referenzzellen indirekt benutzt;

Fig. 13B ist eine programmierbare Schaltung für vorgespanntes Lesen der Master-Referenzzellen gemäß dem alternativen Ausführungsbeispiel;

20 Fig. 13C ist ein Detailschaltkreisdiagramm für die programmierbare Vorspannungsschaltung gemäß Fig. 13B;

Fig. 13D zeigt einen Lesealgorithmus für das Ausführungsbeispiel gemäß Fig. 13A;

25 Fig. 14 veranschaulicht die Lese/Programmier-Datenwege für einen Haufen paralleler Zellen;

Fig. 15 zeigt einen "On-Chip"-Programmier/Verifizieralgorithmus gemäß der vorliegenden Erfindung;

30 Fig. 16 ist ein Schaltkreisdiagramm für die Vergleichsschaltung gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

35 Fig. 17 ist ein Schaltkreisdiagramm für die Programmierschaltung mit Sperren gemäß einem Ausführungsbeispiel der vorliegenden Erfindung;

Tabelle 1 und 2 zählen typische Beispiele von Betriebsspannungen für die EEPROM-Zelle der vorliegenden Erfindung auf.

40 Beschreibung der bevorzugten Ausführungsbeispiele

Es gibt für die Herstellung eines Speicherfeldes viele spezielle EPROM-, EEPROM-Halbleiter-IC-Konstruktionen, bei denen die verschiedenen Aspekte der vorliegenden Erfindung mit Vorteil verwirklicht werden.

5 EEPROM-Zelle mit "geteiltem Kanal"

Ein bevorzugter EEPROM-Aufbau ist in den Querschnittsansichten der integrierten Schaltung gemäß Fig. 1 und 2 insgesamt dargestellt. Um diese bevorzugte Konstruktion kurz zu beschreiben, werden zwei Speicherzellen 11 und 13 auf einem geringfügig p-dotierten Substrat 15 gebildet. Ein stark n-dotierter implantierter Bereich 17 zwischen den Zellen 11 und 13 dient als Drain für die Zelle 11 und als Source für die Zelle 13. In ähnlicher Weise ist ein weiterer implantierter n-dotierter Bereich 19 Source für die Zelle 11 und Drain für eine benachbarte Zelle, und ähnlich ist es mit einem weiteren n-dotierten Bereich 21.

15 Jede der Speicherzellen 11 und 13 enthält entsprechende, leitfähige schwimmende Gates 23 und 25, die insgesamt aus einem Polysiliziummaterial gemacht sind. Jedes dieser schwimmenden Gates ist von einem dielektrischen Material umgeben, um sie voneinander sowie von jeglichen weiteren leitfähigen Elementen der Konstruktion zu isolieren. Ein Steuergate 27 erstreckt sich über beide Zellen 11 und 13 in solcher Weise, daß es gegenüber den schwimmenden Gates 20 und dem Substrat selbst isoliert ist. Wie Fig. 2 zeigt, sind zusätzlich leitfähige Streifen 29 und 31 vorgesehen, die gegenüber einander und sonstigen leitfähigen Elementen der Konstruktion isoliert sind und als Löschgates dienen. Ein Paar solcher Löschgates umgibt das schwimmende Gate jeder Speicherzelle und sind durch eine dielektrische Löschsicht von ihm getrennt. Die Zellen sind durch dicke Feldoxidbereiche, wie die Bereiche 33, 35 und 37 im Querschnitt der 25 Fig. 1 und Bereiche 39 und 41 in der Ansicht gemäß Fig. 2 isoliert.

Die Speicherzelle wird programmiert durch die Übertragung von Elektronen vom Substrat 15 in ein schwimmendes Gate, beispielsweise das schwimmende Gate 25 der Speicherzelle 13. Die Ladung des schwimmenden Gates 25 wird durch Elektronen erhöht, die sich durch das Dielektrikum hinweg aus einem stark p-dotierten Bereich 43 auf das schwimmende Gate bewegen. Ladung wird vom schwimmenden Gate durch das Dielektrikum zwischen ihm und den Löschgates 29 und 31 entfernt. Diese bevorzugte EEPROM-Konstruktion und ein Verfahren zum Herstellen derselben ist im einzelnen in US-A-5 070 032 beschrieben.

35 Bei der in Fig. 1 und 2 dargestellten EEPROM-Konstruktion handelt es sich um einen Typ mit "geteiltem Kanal". Jede Zelle läßt sich als ein Verbundtransistor betrachten, der aus zwei Transistoren T1 und T2 in Reihenschaltung besteht, wie Fig. 3 zeigt. Der T1-Transistor 11a ist längs der Länge L1 des Kanals der Zelle 11 gemäß Fig. 1 ausgebildet. Er hat eine veränderliche Schwellenspannung  $V_{T1}$ . In Reihe mit dem T1-Transistor 11a liegt der T2-Transistor 11b, der in 40 einem Abschnitt des Kanals L2 ausgebildet ist. Er hat eine feste Schwellenspannung  $V_{T2}$  von etwa 1 V. Elemente des Ersatzschaltkreises gemäß Fig. 3 sind mit den gleichen Bezugsziffern wie für die entsprechenden Teile in Fig. 1 und 2, unter Hinzufügung eines Apostroph (') bezeichnet.

Wie am besten aus dem Ersatzschaltkreis gemäß Fig. 3 zu entnehmen ist, beeinflußt der Ladungsspeigel am schwimmenden Gate 23' des T1 einer EEPROM-Zelle die Schwellenspannung  $V_{T1}$  des T1-Transistors 11a bei Betätigung mit dem Steuergate 27'. So läßt sich eine Anzahl von Speicherzuständen in einer Zelle entsprechend gut definierten Schwellenspannungen bestimmen, die durch geeignete, dem schwimmenden Gate auferlegte Ladungsmengen in die Zelle programmiert werden. Das Programmieren erfolgt durch Anlegen geeigneter Spannungen an das Steuergate 27' ebenso wie an Drain 17' und Source 19' der Zelle über eine gewisse Zeitspanne hinweg.

10 Adressierbares Flash-EEPROM-Feld

Die verschiedenen Aspekte der vorliegenden Erfindung werden typischerweise auf ein Feld von Flash-EEPROM-Zellen in einem integrierten Schaltungschip angewandt. In Fig. 4 ist ein Feld individuell adressierbarer EEPROM-Zellen 60 schematisch dargestellt. Jede Zelle ist gleichwertig der einen in Fig. 3 gezeigten und hat ein Steuergate, Source und Drain, sowie ein Löschgate. Die Vielzahl einzelner Speicherzellen sind in Reihen und Spalten organisiert. Jede Zelle wird durch gleichzeitiges wahlweises Erregen ihrer Zeile und Spalte adressiert. So gehört beispielsweise zu einer Spalte 62 eine erste Speicherzelle 63, eine benachbarte zweite Speicherzelle 65 und so weiter. Zu einer zweiten Spalte 72 gehören Speicherzellen 73, 75 und so weiter. Die Zellen 63 und 73 befinden sich in einer Reihe 76, die Zellen 65 und 71 in einer weiteren, benachbarten Reihe und so weiter.

Längs jeder Reihe ist eine Wortleitung mit allen Steuergates der Zellen in der Reihe verbunden. So hat zum Beispiel die Reihe 76 die Wortleitung 77 und die nächste Reihe die Wortleitung 79. Ein Reihendekodierer 81 verbindet wahlweise die Steuergatespannung  $V_{CG}$  auf einer Eingabeleitung 83 mit allen Steuergates längs einer gewählten Wortleitung für eine Reihe.

Längs jeder Spalte ist jeweils die Source aller Zellen durch eine Sourceleitung, beispielsweise 91 und der Drain aller Zellen durch eine Drainleitung, beispielsweise 93 verbunden. Da die Zellen längs einer Reihe jeweils durch ihre Source und Drain in Reihe geschaltet sind, ist der Drain einer Zelle gleichfalls die Source der benachbarten Zelle. So ist die Leitung 93 Drainleitung für die Spalte 62 und gleichzeitig Sourceleitung für die Spalte 72. Ein Spaltendekodierer 101 verbindet wahlweise die Sourcespannung  $V_S$  auf einer Eingabeleitung 103 mit allen Sources und verbindet die Drainspannung  $V_D$  auf einer Eingabeleitung 105 mit allen Drains längs einer gewählten Spalte.

Jede Zelle wird durch die Reihe und Spalte, in der sie sich befindet, adressiert. Wenn zum Beispiel die Zelle 75 zum Programmieren oder Lesen adressiert wird, müssen an das Steuergate, an Source und Drain der Zelle entsprechende Programmier- oder Lesespannungen angelegt werden. Eine Adresse auf dem internen Adressenbus 111 wird zum Dekodieren des Reihendekodierers 81 benutzt, um  $V_{CG}$  auf die mit dem Steuergate der Zelle 75 verbundene Wortleitung 79 zu schalten. Die gleiche Adresse wird zum Dekodieren des Spaltendekodierers 101 benutzt, um  $V_S$  auf die Sourceleitung 93 und  $V_D$  auf die Drainleitung 95 zu schalten, die mit Source bzw. Drain der Zelle 75 verbunden sind.

Ein Ausführungsbeispiel der vorliegenden Erfindung, welches in einem späteren Abschnitt im einzelnen offenbart wird, besteht in der Verwirklichung des Programmierens und Lesens einer Vielzahl paralleler Speicherzellen. Um eine Vielzahl Spalten gleichzeitig auszuwählen, steuert der Spaltendekodierer seinerseits das Schalten eines Sourcingmultiplexers 107 und eines Drainmultiplexers 109. Auf diese Weise können in der gewählten Vielzahl von Spalten jeweils die Sourceleitungen und die Drainleitungen zum Anschluß an  $V_S$  bzw.  $V_D$  zugänglich gemacht werden.

Der Zugang zum Löschgate jeder Zelle ist ähnlich dem zum Steuergate. Bei einer Verwirklichung ist eine Löschleitung, beispielsweise 113 oder 115 oder 117 mit dem Löschgate jeder Zelle einer Reihe verbunden. Ein Löschdekomodierer 119 dekodiert eine Adresse auf dem internen Adressenbus 111 und schaltet wahlweise die Löschspannung  $V_{EG}$  der Eingabeleitung 121 auf eine Löschleitung. Das ermöglicht es jeder Zellenreihe, beispielsweise der Reihe 76 unabhängig adressiert zu werden, die dabei durch die richtigen Spannungen, welche über eine Löschleitung 113 an ihre Löschgates angelegt werden, gleichzeitig (Flash) gelöscht werden. In diesem Fall besteht die Flash-Zelle aus einer Reihe Speicherzellen. Allerdings sind auch andere Verwirklichungen der Flash-Zelle möglich, und in den meisten Anwendungsfällen ist für eine gleichzeitige Löschung vieler Zellenreihen zur gleichen Zeit gesorgt.

20 Flash-EEPROM-System

Das adressierbare EEPROM-Feld 60 in Fig. 4 ist ein Teil des größeren Flash-EEPROM-Systems mit Mehrfachzuständen gemäß der vorliegenden Erfindung, wie es in Fig. 5 veranschaulicht ist. In dem größeren System wird ein integrierter EEPROM-Schaltungsschip 130 von einem Kontroller 140 über eine Schnittstelle 150 gesteuert. Der Kontroller 140 selbst steht in Verbindung mit einer zentralen Mikroprozessoreinheit 160.

Der EEPROM-Chip 130 weist das adressierbare EEPROM-Feld 60, eine serielle Protokolllogik 170, örtliche Stromsteuerschaltungen 180 sowie verschiedene Programmier- und Leseschaltungen 190, 200, 210, 220, 230 und 240 auf.

Der Kontroller 140 steuert die Arbeit des EEPROM-Chips 130 durch Zufuhr der entsprechenden Spannungen, Steuerungen und Zeitbestimmung. Die Tabellen 1 und 2 zeigen typische Beispiele von Spannungsbedingungen für die verschiedenen Betriebsweisen der EEPROM-Zelle. Das adressierbare EEPROM-Feld 60 kann entweder direkt vom Kontroller 140 mit Strom versorgt werden oder, wie in Fig. 5 gezeigt, weiter auf dem Chip durch die örtliche Stromsteuerung 180 geregelt werden. Steuer- und Datenverbindungen zwischen dem Kontroller 140 und dem Chip 130 erfolgen über die serielle EIN-Leitung 251 und die serielle AUS-Leitung 253. Taktzeitsteuerung wird vom Kontroller über Leitung 255 geboten.

Bei einer typischen Operation des EEPROM-Chips 130 sendet der Kontroller 140 einen seriellen Signalstrom über die serielle EIN-Leitung 251 an den Chip 130. Die Signale, die Steuerungs-, Daten-, Adressen- und Zeitinformation enthalten, werden von der seriellen Protokolllogik 170 sortiert. In angemessener Zeitfolge gibt die Logik 170 verschiedene Steuersignale 257 ab, um

die verschiedenen, auf dem Chip 130 vorgesehenen Schaltkreise zu steuern. Sie sendet außerdem eine Adresse über den internen Adressenbus 111, um auf die adressierte Zelle Spannungen zu schalten, die vom Kontroller ausgegeben werden. Wenn es sich bei der Operation um Programmieren handelt, werden in der Zwischenzeit die Daten zur Programmierung der adressierten Zelle zwischengespeichert, indem sie über eine serielle Datenleitung 259 einem Satz Lese/ Programmierzwischenspeicher und Schieberegistern 190 zugesandt werden.

#### Leseschaltungen und Techniken unter Verwendung von Referenzzellen

10 Für EEPROM-Operationen ist es wesentlich, den Speicherzustand einer Zelle exakt und zuverlässig festzustellen. Das liegt daran, daß alle Grundfunktionen, wie Lesen, Löschen, Verifizieren und Programmverifizieren davon abhängen. Verbesserte und neuartige Leseschaltungen 220 für den EEPROM-Chip 130 und Verfahren gemäß der vorliegenden Erfindung machen ein EEPROM mit Mehrfachzuständen realisierbar.

15 Wie schon im Zusammenhang mit Fig. 3 erörtert, wird die programmierte Schwellenspannung  $V_{T1}$  der Zelle durch die programmierte Ladung festgelegt, die dem schwimmenden Gate 23' vermittelt wird. Allgemein gesagt, steigt oder fällt  $V_{T1}$  mit der Menge negativer Ladung am schwimmenden Gate 23'. Die Ladung kann sogar auf einen positiven Wert (Verarmungsmodus) 20 reduziert werden, wobei  $V_{T1}$  auf unterhalb  $V_{T2}$  abnimmt und sogar negativ wird. Der Maximal- und Minimalwert von  $V_{T1}$  richtet sich nach der dielektrischen Festigkeit des Vorrichtungsmaterials. Die Spanne von  $V_{T1}$  bestimmt ein Schwellenspannungsfenster, in welchem Speicherzustände verwirklicht werden können.

25 US-A-5 095 344 offenbart eine EEPROM-Zelle mit Speicherzuständen, die innerhalb eines maximierten Fensters der Schwellenspannung  $V_{T1}$  definiert sind. Das volle Schwellenspannungsfenster schließt den negativen Bereich der Schwellenspannung zusätzlich zu dem üblichen positiven Bereich ein. Das vergrößerte Fenster bietet mehr Speicherplatz zur Verwirklichung von Mehrfachzuständen in einer EEPROM-Zelle.

30 In den Fig. 6 und 7 ist die Art und Weise dargestellt, in der das Schwellenspannungsfenster für einen Speicher mit 2-Zuständen bzw. eine Speicherzelle mit 4-Zuständen aufgeteilt ist. (Natürlich ist es auch möglich, das Fenster für einen Speicher mit 3-Zuständen oder sogar für ein Kontinuum von Zuständen in einem analogen statt digitalen Speicher aufzuteilen.)

35 Zunächst zeigt in Fig. 6 die durchgezogene Kurve 343  $V_{T1}$  als Funktion der Programmierzeit. Das Schwellenspannungsfenster ist vom Minimal- und Maximalwert von  $V_{T1}$  begrenzt, die etwa wiedergegeben sind durch den Löschen-Zustandspegel 345 bzw. den voll Programmieren-Zustandspegel 347. Der Speicher mit 2-Zuständen ist verwirklicht durch Aufteilen des Fensters 40 in zwei Hälften 346, 348 unter Verwendung eines Trennschwellenpegels 349. So läßt sich die Zelle als in einem Speicherzustand 0 (oder Zustand 1) befindlich betrachten, wenn die Zelle mit einem  $V_{T1}$  innerhalb des Bereichs 346 (oder Bereichs 348) programmiert ist.

Ein typischer Lösch/Programmierzyklus beginnt mit dem Löschen, wodurch die Schwellenspannung der Zelle auf den Löschen-Zustandspiegel 345 reduziert wird. Anschließend wird wiederholt programmiert, um die Schwellenspannung  $V_{T1}$  auf den gewünschten Pegel anzuheben. Statt an die adressierte Zelle kontinuierlich eine feste Zeitlang Programmierspannungen entsprechend dem Zustand, auf den die Zelle programmiert werden soll, anzulegen, wird bevorzugt, Programmierspannungen in wiederholten kurzen Impulsen anzulegen, wobei nach jedem Impuls eine Leseoperation stattfindet, um festzustellen, wann sie auf den gewünschten Schwellenspannungspiegel programmiert wurde, einem Zeitpunkt, an dem das Programmieren endet. Die Programmierspannungen und die Dauer der Impulse sind derartig, daß die Impulse  $V_{T1}$  rasch über die verschiedenen Bereiche vorwärtsbringen, wobei aber jeder Impuls ausreichend fein ist, um über keinen der Bereiche hinauszuschießen. Hierdurch werden Beanspruchungen der Zelle im Zusammenhang mit Spannung und Feld auf ein Minimum eingeschränkt und infolgedessen die Zuverlässigkeit der Zelle verbessert.

In Fig. 7A ist der Fall mit 4-Zuständen dargestellt, wo das Schwellenspannungsfenster durch Trennpegel 352, 354, 356 in vier Bereiche 351, 353, 355 bzw. 357 aufgeteilt ist. Es wird davon ausgegangen, daß sich die Zelle im Zustand "3" oder "2" oder "1" oder "0" befindet, wenn ihre  $V_{T1}$  so programmiert ist, daß sie innerhalb entsprechender Bereiche 351 oder 353 oder 355 oder 357 liegt. Eine Zelle mit 4-Zuständen kann zwei Bits Daten speichern. Somit lassen sich die vier Zustände kodieren als (1,1), (1,0), (0,1) bzw. (0,0).

Allgemein gesagt, muß das Schwellenfenster, wenn jede EEPROM-Zelle K Zustände speichern soll, in K Bereiche mit mindestens  $K-1$  Schwellenpegeln aufgeteilt werden. Es ist also nur ein Trennpegel für eine Speicherzelle mit 2-Zuständen erforderlich, und es sind drei Trennpegel nötig für eine Zelle mit 4-Zuständen.

Im Prinzip kann das Schwellenspannungsfenster in eine große Anzahl von Speicherzuständen aufgeteilt werden. So kann es zum Beispiel für eine EEPROM-Vorrichtung mit einem maximalen Schwellenfenster von 16 V in zweieinhalb Zustände aufgeteilt werden, und zwar jeweils innerhalb eines Intervalls von etwa einem halben Volt. In der Praxis haben bekannte EEPROM-Vorrichtungen nur zwei Zustände, oder ein Bit pro Zelle gespeichert, mit verringriger Zuverlässigkeit und Lebensdauer. Abgesehen davon, daß sie mit einem kleineren Schwellenfenster arbeiten, können bekannte Vorrichtungen auch zwei weitere Probleme nicht lösen, die EEPROM-Vorrichtungen inhärent sind. Beide Probleme haben mit der Ungewißheit der Ladungsmenge im schwimmenden Gate und damit mit der Ungewißheit der in die Zelle programmierten Schwellenspannung  $V_{T1}$  zu tun.

Das erste Problem hat mit der zur Dauerleistung in Beziehung stehenden Beanspruchung zu tun, die die Vorrichtung jedes Mal erleidet, wenn sie einen Lösch/Programmierzyklus durchläuft. Die Dauerleistung einer Flash-EEPROM-Vorrichtung ist ihre Fähigkeit, einer gegebenen Anzahl von Programmier/Löszyklen standzuhalten. Das physikalische Phänomen, welches die Dauerleistung bekannter Flash-EEPROM-Vorrichtungen begrenzt, besteht im Einfangen von Elektronen in den aktiven dielektrischen Filmen der Vorrichtung. Während des Programmierens werden Elektronen aus dem Substrat durch eine dielektrische Grenzfläche in das schwimmende Gate inji-

ziert. Ähnlich werden während des Löschens Elektronen aus dem schwimmenden Gate extrahiert, um das Gate durch eine dielektrische Grenzfläche hindurch zu löschen. In beiden Fällen werden von der dielektrischen Grenzfläche einige der Elektronen eingefangen. Die eingefangenenen Elektronen stehen dem angelegten elektrischen Feld in anschließenden Programmier/Löszyklen entgegen und verursachen damit eine Verschiebung der programmierten  $V_{T1}$  auf einen niedrigeren Wert und der gelöschten  $V_{T1}$  auf einen höheren Wert. Das kann man anhand eines graduellen Schließens im Spannungs-“Fenster” zwischen dem "0"- und dem "1"-Zustand bekannter Vorrichtungen sehen, wie in Fig. 8A gezeigt. Jenseits von etwa  $1 \times 10^4$  Programmier/Löszyklen kann das Schließen des Fensters so schwerwiegend werden, daß die Leseschaltungsanordnung nicht mehr richtig funktioniert. Werden die Zyklen dann fortgesetzt, wird die Vorrichtung schließlich einen katastrophalen Ausfall aufgrund eines Dielektrikumsbruchs erleiden. Das geschieht typischerweise bei zwischen  $1 \times 10^6$  und  $1 \times 10^7$  Zyklen und ist als Eigenversagen der Vorrichtung bekannt. Bei bekannten EEPROM-Vorrichtungen ist es das Schließen des Fensters, das die praktische Dauerleistung auf etwa  $1 \times 10^4$  Programmier/ Löszyklen begrenzt. Dieses Problem ist noch kritischer, wenn ein Speicher mit Mehrfachzuständen verwirklicht wird, da eine noch genauere Plazierung von  $V_{T1}$  gefragt ist.

Ein zweites Problem hat mit dem Halten der Ladung am schwimmenden Gate zu tun. Die Ladung am schwimmenden Gate hat die Tendenz, sich im Verlauf der Zeit durch Leckverluste etwas abzuschwächen. Das verursacht auch eine Verschiebung der Schwellenspannung  $V_{T1}$  auf einen niedrigeren Wert im Verlauf der Zeit. Fig. 8B zeigt die Minderung von  $V_{T1}$  in Funktion der Zeit. Im Verlauf der Lebensdauer der Vorrichtung kann  $V_{T1}$  sich um bis zu 1 V verschieben. In einer Vorrichtung mit Mehrfachzuständen könnte dies den Speicher um einen oder zwei Zustände verlagern.

Die vorliegende Erfindung überwindet diese Probleme und bietet Schaltkreise und Verfahren zum zuverlässigen Programmieren und Lesen der verschiedenen Zustände selbst bei einer Verwirklichung mit Mehrfachzuständen.

Der Speicherzustand einer Zelle kann festgestellt werden durch das Messen der darin programmierten Schwellenspannung  $V_{T1}$ . Gemäß einer Alternative kann der Speicherzustand, wie in US-A-5 095 344 erläutert, zweckmäßigerweise bestimmt werden durch das Messen der unterschiedlichen Leitung im Source-Drain-Strom  $I_{DS}$  für die verschiedenen Zustände. Für das Beispiel mit den 4-Zuständen zeigt Fig. 7A die Aufteilung im Schwellenspannungsfenster. Fig. 7B andererseits zeigt typische Werte von  $I_{DS}$  (durchgezogene Kurven) für die vier Zustände als Funktion der Steuergatespannung  $V_{CG}$ . Wenn  $V_{CG}$  bei 5 V liegt, können die  $I_{DS}$ -Werte für jeden der vier Leitzustände dadurch unterschieden werden, daß mit vier entsprechenden Stromabtastverstärkern parallel abgetastet wird. Jedem Verstärker ist ein entsprechender  $I_{REF}$ -Pegel als Bezugsleitzustand (in Fig. 8 als unterbrochene Kurven gezeigt) zugeordnet. Ebenso wie die Trennschwellenpegel (siehe Fig. 6 und 7A) zum Abgrenzen der unterschiedlichen Bereiche im Schwellenspannungsfenster benutzt werden, werden die  $I_{REF}$ -Pegel benutzt, um das gleiche im entsprechenden Source-Drain-Stromfenster zu tun. Durch einen Vergleich mit den  $I_{REF}$ -Werten kann der Leitzustand der Speicherzelle festgestellt werden. In US-A-5 095 344 wird vorgeschlagen, die gleichen Abtastverstärker und  $I_{REF}$ -Werte sowohl für das Programmieren als

auch das Lesen zu benutzen. Das ergibt einen guten Gleichlauf zwischen den Referenzpegeln (unterbrochene Kurven in Fig. 89) und den programmierten Pegeln (durchgezogene Kurven in Fig. 7B).

- 6 In dem verbesserten Plan der vorliegenden Erfindung werden die  $I_{REF}$ -Werte selbst von den Source-Drain-Strömen eines Satzes EEPROM-Zellen bereitgestellt, die auf dem gleichen Chip bestehen und allein für diesen Zweck reserviert sind. Sie wirken also als Master-Referenzzellen, wobei ihre  $I_{REF}$ -Werte als Referenzpegel für das Lesen und Programmieren aller übrigen EEPROM-Zellen auf dem gleichen Chip benutzt werden. Wenn die gleiche Vorrichtung wie die 10 EEPROM-Zellen benutzt wird, um als Referenzzellen zu dienen, wird ein ausgezeichneter Gleichlauf hinsichtlich Temperatur, Spannung und Prozeßschwankungen erzielt. Ferner wird das Problem mit dem Halten der Ladung, was bei der Verwirklichung mit Mehrfachzuständen wichtig ist, erleichtert.
- 15 Unter Hinweis auf Fig. 9A ist eine derartige Master-Referenzzelle 400 mit ihren Programmier- und Lesewegen gezeigt. Das Referenzzellenlösch- und Programmiermodul 410 dient zum Programmieren oder Neuprogrammieren jeder derartigen Referenzzelle 400. Das Modul 410 schließt Programmier- und Löschschaltungen 411 mit einem an den Drain der Master-Referenzzelle 400 angeschlossenen Programmierweg 413 ein. Die Schaltungen 411 werden mittels 20 Adressen initiiert, die vom internen Bus 111 durch einen Programmierdekodierer 415 bzw. einen Löschdekodierer 417 dekodiert werden. Dementsprechend werden Programmierspannungen oder Löschspannungen wahlweise jeder Referenzzelle, beispielsweise der Zelle 400 zugeführt. Auf diese Weise kann der Referenzpegel in jeder Referenzzelle unabhängig gesetzt oder neuprogrammiert werden. Es ist üblich, daß der Schwellenpegel jeder Referenzzelle von seiten der 25 Fabrik auf den optimalen Pegel programmiert ist, der für jede hergestellte Charge Chips angemessen ist. Das könnte durch einen Vergleich mit einem externen Standardbezugspegel geschehen. Auch durch Softwaresteuerung hat ein Nutzer die Option, die Referenzschwellenpegel neu zu setzen.
- 30 Sobald die Referenzschwellenspannung  $V_{T1}$  oder der Referenz-Drain-Source-Strom  $I_{REF}$  in jede Referenzzelle 400 programmiert ist, dient diese als Referenz für das Lesen einer adressierten Speicherzelle, beispielsweise der Zelle 420. Die Referenzzelle 400 ist mit einem ersten Schenkel 403 eines Stromabtastverstärkers 410 über einen getakteten Schalter 413 verbunden. Ein zweiter Schenkel 415 des Verstärkers ist im wesentlichen mit der adressierten Speicherzelle 420 35 verbunden, deren programmierte Leitzustand festgestellt werden soll. Wenn die Zelle 420 gelesen werden soll, wird ein Schalter 421 durch ein Steuersignal READ aktiviert, so daß der Drain der Zelle mit dem zweiten Schenkel 415 verbunden wird. Der Abtastverstärker 410 liefert Spannung über  $V_{CC}$  an die Drains sowohl der Master-Referenzzelle 400 als auch der adressierten Zelle 420. Bei dem bevorzugten Ausführungsbeispiel hat der Verstärker eine Stromspiegel- 40 konfiguration, so daß jeglicher Unterschied in Strömen durch die beiden Schenkel 403 und 415 dazu führt, daß die Spannung im zweiten Schenkel 415 nach oben zu  $V_{CC}$  oder nach unten zu  $V_S$  gezogen wird. Damit ist der Knoten am zweiten Schenkel 415 jeweils HIGH (oder LOW), wenn der Source-Drain-Strom  $I_{DS}$  für die adressierte Zelle 420 weniger (oder mehr) ist als  $I_{REF}$  durch die Master-Referenzzelle 400. Zum geeigneten, durch einen getakteten Schalter 423

bestimmten Zeitpunkt kann das erfaßte Ergebnis am zweiten Schenkel 415 von einem Zwischenspeicher 425 gehalten und an einer Ausgabeleitung 427 verfügbar gemacht werden. Wenn  $I_{DS}$  kleiner ist als  $I_{REF}$ , erscheint an der Ausgabeleitung 427 HIGH, und die adressierte Zeile 420 wird als im gleichen Leitzustand befindlich betrachtet wie die Master-Referenzzeile 400.

Bei dem bevorzugten Ausführungsbeispiel ist zwischen dem zweiten Schenkel 415 und dem Drain 431 der adressierten Zelle 420 auch eine Schaltung 430 zur Spannungsbegrenzung und für raschen Pull-up eingesetzt. Die Schaltung 430 dient dazu, die Drainspannung  $V_D$  auf maximal 1,5 V - 2,0 V zu halten, wenn sie im Fall von niedrigerem  $I_{DS}$  auflädt. Sie verhindert auch, daß  $V_D$  im Fall von höherem  $I_{DS}$  einen zu tiefen Pull-down erfährt.

Wenn jede Speicherzelle K Zustände speichern soll, müssen insgesamt mindestens K-1 oder vorzugsweise K Bezugspegel vorgesehen werden. Bei einem Ausführungsbeispiel wird die adressierte Zelle mit den K Bezugszellen unter Verwendung von k parallelen Abtastverstärkern verglichen. Das ist von Vorteil für den Fall mit 2-Zuständen wegen der Geschwindigkeit, kann aber den verfügbaren Strom zu dünn ausbreiten, als daß er im Fall mit Mehrfachzuständen ordnungsgemäß erfaßt werden könnte. Für den Fall mit Mehrfachzuständen ist deshalb vorzuziehen, die adressierte Zelle mit K Bezugszellen, und zwar eine nach der anderen zu vergleichen.

Fig. 9B zeigt deutlicher die Lesekonfiguration für Mehrfachzustände. Die K Referenzzellen, wie 431, 433, 435 sind mit dem Abtastverstärker 440 über den ersten Schenkel 441 des Verstärkers verbunden. Die Verbindung ist durch getaktete Schalter, beispielsweise 451, 453 bzw. 455 zeitmultiplexiert. Der zweite Schenkel 457 des Abtastverstärkers ist mit der adressierten Zelle verbunden, wie in Fig. 9A. Das Abtastsignal am zweiten Schenkel 457 wird durch getaktete Schalter, beispielsweise 461, 463, 465 zeitwahliweise auf solche Zwischenspeicher 471, 473, 475 gegeben.

Die Fig. 9C(1)-9C(8) veranschaulichen den zeitlichen Ablauf für das Lesen von Mehrfachzuständen. Wenn das Signal READ auf HIGH geht, wird ein Schalter 421 aktiviert und die adressierte Speicherzelle mit dem zweiten Schenkel 457 des Abtastverstärkers 440 verbunden (Fig. 9C(1)). Die Zeitbestimmung der Takte ist in den Fig. 9C(2)-9C(4) gegeben. Somit vergleicht der Abtastverstärker bei jedem Taktignal der Reihe nach die adressierte Zelle mit jeder der Referenzzellen und behält jedes der Ergebnisse ein. Die einbehaltenen Ausgaben des Abtastverstärkers sind in den Fig. 9C(5)-9C(7) gegeben. Wenn alle K Ausgabezustände des Abtastverstärkers 440 einbehalten sind, werden sie von einem K-L Dekodierer 480 ( $2^L \geq K$ ) (Fig. 9C(8)) zu L binären Bits kodiert.

Die mehrfachen Schwellenpegel werden also von einem Satz Speicherzellen geliefert, der als Master-Referenzzellen dient. Die Master-Referenzzellen sind unabhängig und von außen lösbar und programmierbar, und zwar entweder vom Hersteller der Vorrichtung oder vom Benutzer. Dieses Merkmal bietet die größtmögliche Flexibilität, indem es ein individuelles Einstellen der Trennschwellen innerhalb des Schwellenfensters der Vorrichtung zu jeder beliebigen Zeit erlaubt. Da es sich um die gleiche Vorrichtung handelt wie die Speicherzellen, haben die

Bezugszellen einen engen Gleichlauf mit den gleichen Schwankungen aufgrund des Herstellungsprozesses, der Betriebsbedingungen sowie Schwierigkeiten im Halten von Ladung. Die unabhängige Programmierbarkeit jedes Schwellenpegels nach Wunsch erlaubt eine Optimierung und Feinabstimmung der Aufteilung des Schwellenfensters, um einen Speicher mit Mehrfachzuständen lebensfähig zu machen. Ferner erlaubt sie eine Konfiguration nach der Herstellung zu einem Speicher mit entweder 2-Zuständen oder mit Mehrfachzuständen anhand der gleichen Vorrichtung, je nach Bedarf des Benutzers oder Merkmalen der Vorrichtung im betreffenden Zeitpunkt.

Ein weiteres Ausführungsbeispiel der vorliegenden Erfindung bietet eine verbesserte Mehrfachzustandabtastung einer adressierten Speicherzelle. Wie im Zusammenhang mit einem früheren Ausführungsbeispiel zum Abtasten eines Mehrfachzustandspeichers erörtert, ist es vorzuziehen, den Leistungsstrom der Zelle mit allen Referenzleistungsstrompegeln (Schwellenpegeln) gleichzeitig oder parallel zu vergleichen. Eine 4-Zustandspeicherzelle hat zum Beispiel mindestens drei Referenzstrompegel zum Abgrenzen der vier Zustände. Ein paralleles Abtasten des Zustands der Zelle bedeutet einen gleichzeitigen Vergleich des Leistungsstroms der Zelle  $I_{CELL}$  gegenüber jedem der drei Referenzstrompegel. Dies ist schneller als ein Vergleich mit jedem der drei Referenzleistungspegel der Reihe nach. Allerdings würde bei dem früher beschriebenen einfacheren Ausführungsbeispiel der Leistungsstrom der adressierten Zelle durch das Aufteilen in drei Zweige, nämlich einen für jeden Referenzpegelvergleich verdünnt. Folglich kann sich eine einfache Verwirklichung gleichzeitiger oder paralleler Mehrfachzustandabtastung durch den erforderlichen Rauschabstand des Abtastsystems verbieten, insbesondere wenn viele Zustände betroffen sind.

Ein weiteres wichtiges Merkmal der vorliegenden Erfindung dient der Überwindung der Schwierigkeiten hinsichtlich der Dauerleistung. Wie zuvor erläutert, hängt die Lösch-Programmier- und Lesekennlinie jeder Speicherzelle von der kumulierten Beanspruchung ab, die während der Anzahl von Programmier/ Löschzyklen erlitten wurde, die die Zelle durchlaufen hat. Insgesamt unterliegen die Speicherzellen viel mehr Programmier/Löschenzyklen als die Master-Referenzzellen. Die anfangs optimierten Referenzpegel werden dadurch schließlich fehljustiert, so daß sie Lesefehler verursachen. Der vorliegende, grundlegende erfinderische Gedanke besteht darin, die Referenzpegel auch den gleichen periodischen Durchlauf reflektieren zu lassen, den die Speicherzellen erleiden. Das wird erreicht durch die Verwirklichung lokaler Referenzzellen zusätzlich zu den Master-Referenzzellen. Die lokalen Referenzzellen werden den gleichen periodischen Programmier/Löschenzyklen unterworfen wie die Speicherzellen. Nach jeder Löschoperation werden die Referenzpegel in den Master-Referenzzellen neu in den entsprechenden Satz lokaler Referenzzellen kopiert. Dann werden Speicherzellen mit Bezug auf die Referenzpegel der in engem Gleichlauf befindlichen lokalen Referenzzellen gelesen. Auf diese Weise wird das Abweichen der Zellkennlinie nach jedem Programmier/Löschenzyklus automatisch ausgeglichen. Deshalb wird die richtige Unterteilung des Umformschwellenfensters beibehalten, so daß die Speicherzustände selbst nach vielen Zyklen korrekt gelesen werden können.

Die Verwirklichung mit dem Referenzieren lokaler Zellen für Flash-EEPROM ist in Fig. 10 dargestellt. In dem Flash-EEPROM-Feld 60 (Fig. 4) wird jede Gruppe Speicherzellen, die kollektiv

gelöscht oder programmiert werden, als ein Sektor bezeichnet. Der Ausdruck "Flash-Sektor" ist analog dem in Magnetplat-tenspeichervorrichtungen benutzten Ausdruck "Sektor", und sie werden hier austauschbar benutzt. Das EEPROM-Feld ist in Flash-Sektoren gruppiert, beispielsweise 501, 503 und 505. Während alle Speicherzellen in einem Flash-Sektor den gleichen periodischen Durchlauf erfahren, können verschiedene Flash-Sektoren unterschiedliche periodische Durchläufe erfahren. Um für jeden Flash-Sektor Gleichlauf sicherzustellen, wird ein Satz Speicherzellen in jedem Flash-Sektor zur Benutzung als lokale Referenzzellen reserviert. Nachdem beispielsweise der Flash-Sektor 503 gelöscht wurde, werden die Referenzpegel in den Master-Referenzzellen 507 in die zum Flash-Sektor 503 gehörenden lokalen Referenzzellen neuprogrammiert. Bis zum nächsten Löschzyklus setzen die Leseschaltungen 513 das Lesen der Speicherzellen innerhalb des Flash-Sektors 503 mit Bezug auf die neuprogrammierten Referenzpegel fort.

Fig. 11(1)-11(7) illustriert den Algorithmus zum Neuprogrammieren der Referenzzellen eines Sektors. Insbesondere beziehen sich die Fig. 11(1)-11(3) auf das Löschen der lokalen Referenzzellen des Sektors auf ihre "gelöschten Zustände". So wird in Fig. 11(1) ein Löschspannungsimpuls an alle Speicherzellen des Sektors, einschließlich der lokalen Referenzzellen angelegt. In Fig. 11(2) werden dann alle lokalen Referenzzellen mit Bezug auf die Master-Referenzzellen gelesen, um zu verifizieren, ob sie alle auf den "gelöschten Zustand" gelöscht worden sind.

Solange noch eine Zelle als anderweitig festgestellt wird, wird ein weiterer Löschspannungsimpuls an alle Zellen angelegt. Dieser Prozeß wird wiederholt, bis von allen lokalen Referenzzellen im Sektor mittels Überprüfung festgestellt wurde, daß sie sich im "gelöschten" Zustand befinden (Fig. 11(3)).

Fig. 11(4)-11(7) beziehen sich auf das Programmieren der lokalen Referenzzellen im Sektor. Nachdem der "gelöschte" Zustand aller lokalen Referenzzellen im Sektor verifiziert wurde, wird ein Programmierspannungsimpuls in Fig. 11(4) nur an alle lokalen Referenzzellen angelegt. Dem folgt in Fig. 11(5) das Lesen der lokalen Referenzzellen mit Bezug auf die Master-Referenzzellen, um zu verifizieren, ob jede einzelne der lokalen Referenzzellen auf den gleichen Zustand programmiert wurde wie die entsprechende Master-Referenzzelle. Für jene lokalen Referenzzellen, für die das nicht verifiziert wurde, wird ein weiterer Programmierspannungsimpuls wahlweise nur an sie allein angelegt (Fig. 11(6)). Dieser Prozeß wird wiederholt, bis die Verifizierung ergibt, daß alle lokalen Referenzzellen korrekt auf die verschiedenen Trennschwellenpegel im Schwellenfenster programmiert sind (Fig. 11(7)).

Sobald die lokalen Referenzzellen im Sektor neuprogrammiert worden sind, werden sie direkt oder indirekt zum Löschverifizieren, Programmierverifizieren oder Lesen der adressierten Speicherzellen des Sektors benutzt.

Fig. 12A veranschaulicht ein Ausführungsbeispiel, bei dem die lokalen Referenzzellen unmittelbar zum Lesen oder Programmier/ Löschverifizieren der Speicherzellen des Sektors benutzt werden. Während dieser Operationen wird also ein paralleles Schalterpaar 521 durch ein READ-Signal aktiviert, und der Abtastverstärker 440 liest die adressierten Speicherzellen 523 des Sektors bezüglich jeder der lokalen Referenzzellen 525 des Sektors. Während des Program-

mier/Löschverifizierens der lokalen Referenzzellen (wie in Fig. 11 dargestellt) ermöglicht ein weiteres paralleles Schalterpaar 527 das Lesen der lokalen Referenzzellen 525 bezüglich der Master-Referenzzellen 529.

5 Fig. 12B veranschaulicht den Algorithmus zur direkten Benutzung der lokalen Referenzzellen zum Lesen oder Programmier/ Löschverifizieren der adressierten Speicherzellen des Sektors.

10 Fig. 13A stellt ein alternatives Ausführungsbeispiel dar, bei dem die lokalen Referenzzellen zum Lesen der adressierten Speicherzellen indirekt benutzt werden. Zunächst werden die Master-Referenzzellen gelöscht und programmiert, und zwar jede auf eine der gewünschten mehrfachen Trennschwellen innerhalb des Schwellenfensters. Unter Benutzung dieser Master-Referenzschwellen werden die lokalen Referenzzellen innerhalb eines gelöschten Zellensektors jeweils auf eine der gleichen gewünschten mehrfachen Trennschwellen programmiert. Als nächstes werden die adressierten Zellen im Sektor mit den gewünschten Daten programmiert (beschrieben). Die 15 Lesesequenz für die adressierten Zellen im Sektor beinhaltet dann die in Fig. 13D dargestellten Schritte.

20 Zunächst wird jede der lokalen Referenzzellen 525 bezüglich der entsprechenden Master-Referenzzelle 531 gelesen. Dies geschieht durch ein aktivierendes READ-1-Signal an einen Schalter 533, der die lokalen Referenzzellen 525 mit dem zweiten Schenkel 457 des Abtastverstärkers 440 verbindet, wobei die Master-Referenz 531 mit dem ersten Schenkel 441 des gleichen Verstärkers verbunden ist. Jeder Master-Referenzzelle zugeordnete Hilfsstromquellenschaltungen werden nunmehr benutzt, um den Strom durch den ersten Schenkel 441 des Abtastverstärkers zur Anpassung an den Strom im zweiten Schenkel 457 optimal vorzuspannen. Wenn 25 die Vorspannungseinstelloperation für alle Trennschwellenpegel vollendet ist, werden die adressierten Zellen im Sektor mit Bezug auf die hinsichtlich der Vorspannung eingestellten Master-Referenzzellen gelesen. Dies geschieht mittels eines deaktivierenden READ-1 an 533 und eines aktivierenden READ-Signals an Schalter 535. Der Vorteil dieses Ansatzes besteht darin, daß jegliche Schwankungen im  $V_{CC}$ , in der Temperatur, Zyklusermüdung oder sonstigen Auswirkungen, die im Verlauf der Zeit Schwellenabweichungen zwischen den Master-Referenzzellen und den adressierten Zellen verursachen können, vor dem Lesen eliminiert wird, da die lokalen Referenzzellen (die Schwellenabweichungen der adressierten Zellen im Gleichlauf verfolgen) benutzt werden, um die Trennschwellen der Master-Referenzzellen wirksam neu einzustellen. Zum Beispiel erlaubt es dieser Plan, die adressierten Zellen zu programmieren, wenn die Master-35 Referenzzellen mit  $V_{CC}=5,5$  V betrieben werden und anschließend die adressierten Zellen zu lesen, wenn die Master-Referenzzellen mit  $V_{CC}=4,5$  V versorgt werden. Die Differenz von 1 Volt im  $V_{CC}$ -Wert, die normalerweise eine Änderung im Wert der Trennschwellen verursachen würde, wird dadurch neutralisiert, daß die lokalen Referenzzellen zur Vorspannungseinstellung der Master-Referenzzellen benutzt werden, um dieser Änderung im Zeitpunkt des Lesens entgegenzuwirken. 40

Die Fig. 13B und 13C zeigen mehr im einzelnen ein Ausführungsbeispiel der Stromvorspannungsschaltungen, beispielsweise 541, 543, 545 für die Master-Referenzzellen 551, 553, 555. Jede Vorspannungsschaltung wirkt als Stromnebenschluß für den Strom in der Master-Refe-

renzzelle. So hat beispielsweise der Schaltkreis 541 einen Abgriff am Drain der Master-Referenzzelle 551 über die Leitung 561. Er modifiziert den Strom in einer Leitung 562 an den Abtastverstärker (erster Schenkel) entweder durch Sourcestrom von  $V_{CC}$  oder Drainstrom an  $V_{SS}$ . Im zuerst genannten Fall wird der Strom in der Leitung 562 vermindert, und es ist anders für den zuletzt genannten Fall. Da für die Master-Referenz 551 Vorspannen bewerkstelligt wird, kann jegliche Ungleichheit der Ströme in den beiden Schenkeln des Abtastverstärkers vom Chip nach außen abgegeben werden. Das wird von dem Kontroller (siehe Fig. 5) festgestellt, der seinerseits die Vorspannschaltung 451 über den internen Adressenbus 111 programmiert, um Strom in der Leitung 562 zu subtrahieren oder zu addieren, damit der der örtlichen Referenz ausgeglichen wird.

Fig. 13C veranschaulicht ein Ausführungsbeispiel der Vorspannschaltung, beispielsweise des Schaltkreises 541. Eine Bank paralleler Transistoren, beispielsweise 571, 573, 575 ist jeweils mit ihrem Drain an  $V_{CC}$  und mit ihrer Source über Schalter, beispielsweise 581, 583, 585 an die Leitung 561 angeschlossen. Durch wahlweises Aktivieren der Schalter kann eine unterschiedliche Anzahl Transistoren benutzt werden, um verschiedene Mengen Strom von der Leitung 562 zu subtrahieren. Ähnlich ist eine weitere Bank paralleler Transistoren, beispielsweise 591, 593, 595 jeweils mit ihrer Source an  $V_{SS}$  angeschlossen und mit ihrem Drain über Schalter, beispielsweise 601, 603, 605 an die Leitung 561. Durch wahlweises Aktivieren der Schalter kann eine unterschiedliche Anzahl Transistoren benutzt werden, um verschiedene Mengen Strom zur Leitung 562 zu addieren. Zum Adressendekodieren vom internen Adressenbus 111 wird ein Dekodierer 609 benutzt, um die Schalter wahlweise zu aktivieren. Die Aktivierungssignale sind in Zwischenspeichern 611, 613 gespeichert. Auf diese Weise werden bei jedem Lesen eines Sektors die Master-Referenzzellen in bezug auf die lokalen Referenzzellen neu vorgespannt und zum Lesen der Speicherzellen im Sektor benutzt.

Die Fig. 13D(1)-13D(4) stellen den Lesealgorithmus für das alternative Ausführungsbeispiel dar. Der Sektor muß zuvor seine lokalen Referenzzellen in bezug auf die Master-Referenzzellen programmiert und verifiziert bekommen haben (Fig. 13D(1)). Dementsprechend wird dann jede der Master-Referenzzellen bezüglich der lokalen Referenzzellen gelesen (Fig. 13D(2)). Die Master-Referenzzellen werden vorgespannt, um den Strom auf den der entsprechenden lokalen Referenzzellen abzugleichen (Fig. 13D(3)). Anschließend werden die Speicherzellen im Sektor mit Bezug auf die vorgespannten Master-Referenzzellen gelesen (Fig. 13D(4)).

Die beschriebenen Leseschaltkreise und die Operation wird auch beim Programmieren und Löschen der Speicherzellen angewandt, insbesondere im Verifizierungsteil des Vorganges. Wie zuvor beschrieben, erfolgt das Programmieren in kleinen Schritten, wobei der programmierte Zustand zwischendurch abgelesen wird, um zu verifizieren, ob der gewünschte Zustand erreicht worden ist. Sobald der programmierte Zustand korrekt verifiziert wurde, hört das Programmieren auf. Ähnlich wird das Löschen in kleinen Schritten vorgenommen, wobei der Löschkzustand zwischendurch gelesen wird, um zu verifizieren, ob der "gelöschte" Zustand erreicht worden ist. Sobald der "gelöschte" Zustand korrekt verifiziert wurde, hört das Löschen auf.

Wie zuvor schon beschrieben, sind nur K-1 Trennschwellenpegel erforderlich, um das Schwellenfenster in K Bereiche aufzuteilen und es damit der Speicherzelle zu erlauben, K Zustände zu speichern. Gemäß einem Aspekt der vorliegenden Erfindung ist es jedoch im Fall des Mehrfachzustandes, wo das Schwellenfenster feiner unterteilt ist, vorzuziehen, K Schwellenpegel für K Zustand zu benutzen. Der zusätzliche Schwellenpegel wird verwendet, um den "gelöschten" Zustand von demjenigen Zustand mit dem niedrigsten Schwellenpegel zu unterscheiden. Das verhindert ein Überlöschen und damit ein Überbeanspruchen der Zelle, da das Löschen sofort aufhört, wenn der "gelöschte" Zustand erreicht ist. Das wahlweise Inhibieren einzelner Zellen zum Löschen gilt nicht für den Flash-EEPROM-Fall, wo jedesmal mindestens ein Sektor gelöscht werden muß. Geeignet sind solche EEPROM-Felder, wo die Speicherzellen individuell zum Löschen adressiert werden können.

Gemäß einem weiteren Merkmal der Erfindung wird eine Speicherzelle, nachdem sie auf den "gelöschten" Zustand gelöscht wurde, geringfügig programmiert, um die Zelle in den Zustand mit dem niedrigsten Schwellenpegel (Erdungszustand) nächst dem "gelöschten" Zustand zu bringen. Dies hat zwei Vorteile. Zunächst sind die Schwellenpegel des Erdungszustandes aller Speicherzellen, da sie begrenzt sind zwischen den beiden gleichen Trennschwellenpegeln, gut definiert und nicht weit verstreut. Das bietet einen gleichmäßigen Ausgangspunkt für das anschließende Programmieren der Zellen. Zweitens erfahren alle Zellen eine gewisse Programmierung, wodurch verhindert wird, daß jene Zellen, bei denen die Tendenz besteht, daß in ihnen der Erdungszustand gespeichert ist, beispielsweise Gleichlauf mit dem Rest hinsichtlich der Programmier/Löszyklen und des Dauerleistungsverhaltens zu verlieren.

#### Programmverifizierung auf dem Chip

Wie zuvor erwähnt, wird vorzugsweise das Programmieren einer EEPROM-Zelle in einen gewünschten Zustand in kleinen Schritten, ausgehend vom "Lös"-Zustand durchgeführt. Nach jedem Programmierschritt wird die Zelle, die gerade programmiert wird, gelesen, um zu prüfen, ob der gewünschte Zustand erreicht worden ist. Ist das nicht der Fall, wird weiteres Programmieren und Verifizieren wiederholt, bis die Prüfung einen solchen Zustand ergibt.

Unter Bezugnahme auf das Systemdiagramm gemäß Fig. 5 steht der EEPROM-Chip 130 unter der Steuerung durch den Kontroller 140. Sie sind seriell in Verbindung durch die SERIAL-IN-Leitung 251 und die SERIAL-OUT-Leitung 253. Bei bekannten EEPROM-Vorrichtungen wird nach jedem Programmierschritt der in der in Programmierung befindlichen Zelle erreichte Zustand gelesen und an den Kontroller 140 oder die CPU 160 zwecks Überprüfung des gewünschten Zustandes zurückgesandt. Diese Vorgehensweise belastet die Geschwindigkeit außerordentlich, insbesondere angesichts der seriellen Verbindung.

In der vorliegenden Erfindung wird die Programmverifizierung optimiert durch das parallele Programmieren eines ganzen Zellenhaufens (typischerweise einige Bytes), gefolgt durch das parallele und auf dem Chip erfolgende Verifizieren. Das parallele Programmieren wird verwirklicht durch eine selektive Programmierschaltung, die das Programmieren jener Zellen in dem Haufen deaktiviert, deren Zustände bereits als richtig überprüft wurden. Dieses Merkmal ist bei

einer Mehrfachzustandsverwirklichung wesentlich, weil einige Zellen ihren gewünschten Zustand früher als andere erreichen und den gewünschten Zustand kontinuierlich weitergeben, wenn man sie nicht anhält. Nachdem der ganze Haufen Zellen korrekt verifiziert wurde, teilt die Logik auf dem Chip diese Tatsache dem Kontroller mit, wodurch das Programmieren des nächsten Zellenhaufens beginnen kann. Auf diese Weise brauchen Daten nicht zwischen dem EEPROM-Chip und dem Kontroller zwischen jedem Programmierschritt hin- und herbewegt zu werden, und die Programmierverifizierungsgeschwindigkeit wird stark gefördert.

Fig. 14 veranschaulicht die Programmier- und Verifizierwege für einen Haufen paralleler  $n$  Zellen. Die gleichen Bezugsziffern sind für entsprechende Module im Systemdiagramm der Fig. 5 benutzt. Das EEPROM-Feld 60 wird auf einmal von  $N$  Zellen adressiert. Zum Beispiel kann  $N$  64 Zellen breit sein. In einem 512 Byte Flash-Sektor, der aus vier Reihen von 1024 Zellen besteht, gibt es 64 Haufen aus 64 Zellen. Der Sourcingmultiplexer 107 verbindet wahlweise die  $N$  Sourcen eines adressierten Zellenhaufens mit der Sourcespannung  $V_S$  auf Leitung 103. Ähnlich macht der Drainmultiplexer 109 wahlweise die  $N$  Drains des Haufens durch einen  $N$ -Kanal Datenweg 105 zugänglich. Während des Programmierens nimmt die Programmschaltung mit Inhibitieren 210 Zugriff auf den Datenweg 105, und während des Lesens, Programmierverifizierens oder Löschverifizierens sind es die Leseschaltkreise 220.

Es sei nochmals auf das Systemdiagramm in Fig. 5 verwiesen. Dort steht die Programmierung unter der Steuerung des Kontrollers 140. Die in den Sektor zu programmierenden Daten werden Haufen für Haufen übermittelt. Zunächst sendet der Kontroller einen ersten Haufen  $N \cdot L$  serieller Datenbits zusammen mit Adressen, Steuer- und Zeitinformationen an den EEPROM-Chip 130.  $L$  ist die Zahl binärer Bits, die pro Speicherzelle kodiert sind. Für eine 2-Zustandzelle ist beispielsweise  $L=1$  und  $L=2$  für eine 4-Zustandzelle. Wenn also  $N=64$  und  $L=2$ , hat der Haufen Datenbits eine Breite von 128 Bits. Die  $N \cdot L$  Datenbits werden in Zwischenspeichern und Schieberegistern 190 gespeichert, wo die seriellen Bits in  $N \cdot L$  parallele Bits umgesetzt werden. Diese Daten sind zur Programmverifizierung im Zusammenhang mit den Leseschaltkreisen 220, dem Bitdekodierer 230, der Vergleichsschaltung 200 und der Programmschaltung mit Inhibitieren 210 erforderlich.

Der Programmalgorithmus für einen Haufen von  $N$  Zellen wird am besten unter Hinweis sowohl auf das Systemdiagramm der Fig. 5 als auch die Fig. 15(1)-15(7) beschrieben, die den Algorithmus selbst darstellen. Wie in einem früheren Abschnitt erwähnt, muß vor dem Programmieren des Sektors der ganze Sektor gelöscht und alle Zellen darin geprüft werden, um zu verifizieren, daß sie sich im "gelöschten" Zustand befinden (Fig. 15(1)). Hierauf folgt in Fig. 15(2) das Programmieren der lokalen Referenzzellen des Sektors (wie in Fig. 11(1)-(3) gezeigt). In Fig. 15(3) werden die  $N \cdot L$  Bits paralleler Daten in Zwischenspeichern 190 einbehalten. In Fig. 15(4) nehmen die Leseschaltkreise 220 Zugriff auf den  $N$ -Kanal Datenweg 105, um die Zustände in dem  $N$  Zellenhaufen zu lesen. Der Lesealgorithmus ist bereits im Zusammenhang mit 12B oder 13D beschrieben worden. Die  $N$ -Zellenablesungen generiert  $N \cdot K$  ( $K$ =Anzahl der Zustände pro Zelle) Ausgabezustände. Diese werden vom Bitdekodierer 230 zu  $N \cdot L$  binären Bits dekodiert. In Fig. 15(5) werden die  $N \cdot L$  Lesebits von der Vergleichsschaltung 200 Bit für Bit mit den  $N \cdot L$  Programmdatenbits von den Zwischenspeichern 190 verglichen. Wenn irgend ein gelesenes Bit

sich nicht mit dem Programmdatenbit vergleichen läßt, wird in Fig. 15(6) ein weiterer Programmierspannungsimpuls von der Programmierschaltung 210 gleichzeitig an den Zellenhaufen angelegt. Allerdings blockiert eine Inhibierschaltung innerhalb der Programmierschaltung 210 wahlweise das Programmieren jener Zellen, deren Bits mit den programmierten Datenbits korrekt verifiziert wurden. Folglich werden nur die nichtverifizierten Zellen jedesmal programmiert. Das Programmieren und Verifizieren wird wiederholt, bis alle Zellen korrekt verifiziert sind in Fig. 15(7).

Fig. 16 zeigt ein Ausführungsbeispiel der Vergleichsschaltung 200 aus Fig. 5 mehr im einzelnen. Die Schaltung 200 weist N Zellenvergleichsmodule auf, beispielsweise 701, 703, und zwar eines für jede der N Zellen im Haufen. In jedem Zellenvergleichsmodul, beispielsweise dem Modul 701 werden die L Lesebits (L = Anzahl der für jede Zelle kodierten binären Bits) Bit für Bit mit den entsprechenden Programmdatenbits verglichen. Das wird von L XOR-Gattern, beispielsweise 711, 713, 715 durchgeführt. Die Ausgabe dieser XOR-Gatter passiert ein NOR-Gatter 717, so daß immer dann eine "1" am Ausgang des NOR-Gatters 717 erscheint, wenn alle L Bits verifiziert sind und ansonsten eine "0" erscheint. Wenn das Steuersignal VERIFY wahr ist, wird dieses Ergebnis in einem Zwischenspeicher 721 einbehalten, so daß das gleiche Ergebnis am Ausgang des NOR-Gatters 717 am Ausgang 725 des Zellenvergleichsmoduls verfügbar ist. Die Vergleichsschaltung 200 führt die Vergleiche von L Bits parallel durch. Die Ausgaben von N Vergleichsmodulen, beispielsweise 725, 727 sind auf einer N-Kanal Ausleitleitung 731 verfügbar, um in die Programmierschaltung mit Inhibieren 210 der Fig. 5 eingegeben zu werden.

Zur gleichen Zeit werden die N Ausgaben, beispielsweise 725, 727 durch ein UND-Gatter 733 geleitet, so daß dessen einzige Ausgabe 735 als Resultat eine "1" ist, wenn alle N Zellen verifiziert sind und eine "0", wenn es anders ist. Unter gleichzeitigem Hinweis auf Fig. 5 wird die einzige Ausgabe 735 benutzt, um dem Kontroller 140 zu signalisieren, daß alle N Zellen im Datenhaufen korrekt verifiziert wurden. Das Signal in der Ausgabe 735 wird während einer VERIFY-Operation durch die SERIAL-OUT-Leitung 253 über ein UND-Gatter 240 gesandt.

Bei "power-up" oder am Ende des Programmier/Verifizierens eines Datenhaufens werden die Ausgänge aller Zellenvergleichsmodule, beispielsweise 725, 727 auf den nichtverifizierten Zustand von "0" zurückgesetzt. Das wird erreicht durch das Anziehen des Knotens 726 an V<sub>SS</sub> (0 V) mittels des RESET-Signals auf der Leitung 727 an einen Transistor 729.

Fig. 17 zeigt in größerer Einzelheit ein Ausführungsbeispiel der Programmierschaltung mit Inhibieren 210 gemäß Fig. 5. Die Programmierschaltung 210 weist N Programm-mit-Inhibiermodule, beispielsweise 801, 803 auf. Wie in Tabelle 1 und 2 dargestellt, muß, um die N Zellen zu programmieren, eine Spannung V<sub>PD</sub> an die Drain jeder einzelnen der N Zellen und eine Spannung V<sub>PG</sub> an die Steuergates angelegt werden. Jedes Programmiermodul, beispielsweise 801 dient zum wahlweisen Durchlassen von V<sub>PD</sub> auf einer Leitung 805 an eine der Drains durch den einen des N-Kanaldataweges 105. Da V<sub>PD</sub> typischerweise etwa 8 V bis 9 V beträgt, was höher ist als V<sub>CC</sub>, kann dieser Wert nicht zum Einschalten des Transistorschalters 807 benutzt werden. Stattdessen wird die höhere Spannung V<sub>CG</sub> (etwa 12 V) zum Aktivieren des Schalters

807 verwendet.  $V_{CG}$  auf der Leitung 801 wird selbst durch ein UND-Gatter aktiviert, wenn sowohl das Programmiersteuersignal PGM auf Leitung 813 wahr und das Signal auf Leitung 731 eine "0" ist. Da das Signal auf der Leitung 731 vom Ausgang des in Fig. 16 gezeigten Zellenvergleichsmoduls 701 kommt, folgt daraus, das  $V_{PD}$  wahlweise an jene Zellen durchgelassen wird, die noch nicht überprüft wurden. Auf diese Weise wird bei jedem Anlegen eines Programmierimpulses dieser nur an jene Zellen angelegt, die ihre beabsichtigten Zustände noch nicht erreicht haben. Dieses selektive Programmiermerkmal ist besonders nötig bei der Verwirklichung des parallelen Programmierens und des Verifizierens auf dem Chip im Fall des Mehrfachzustandes.

10 Variable Steuerung von Spannung an das Steuergate

Das Systemdiagramm gemäß Fig. 5 im Zusammenhang mit den Tabellen 1 und 2 veranschaulicht, wie verschiedene Spannungen an das EEPROM-Feld 60 angelegt werden, um die Grundfunktionen des EEPROMs durchzuführen. Bekannte EEPROM-Vorrichtungen erlauben es der an das Steuergate angelegten Spannung  $V_{CG}$  nur eine von zwei Spannungen anzunehmen, nämlich  $V_{CC}$  oder die höhere Programmierspannung von etwa 12 V.

Bei einem weiteren Aspekt der vorliegenden Erfindung ist es möglich, die an das Steuergate 20 gelieferte Spannung  $V_{CG}$  unabhängig und kontinuierlich über einen breiten Spannungsbereich variabel zu gestalten. Dafür sorgt  $V_{PG}$  vom Kontroller 140. Insbesondere wird  $V_{CG}$  in einer Leitung 83 von  $V_{PG}$  gespeist, und dies wiederum wird vom Kontroller von einer Leitung 901 zugeführt. Tabelle 2 zeigt, daß  $V_{PG}$  verschiedene Spannung unter verschiedenen Funktionen des EEPROMs annimmt.

25 Die Veränderlichkeit von  $V_{CG}$  ist besonders vorteilhaft bei Programmier- und Löschspielraumplänen. Bei der Programmierung mit Spielraum erfolgt das Lesen während der Programmverifizierung mit  $V_{CG}$  auf geringfügig höherer Spannung als der genormten  $V_{CC}$ . Das erleichtert es, die programmierte Schwelle gut in den Zustand hineinzuverlegen, indem die Programmierung 30 mit einer geringfügigen Marge an dem Trennschwellenpegel vorbeigeführt wird. Bei der Löschverifizierung wird die Zelle mit einer etwas reduzierten  $V_{CG}$  überprüft, um die Zelle gut in den "gelöschten" Zustand zu bringen. Ferner kann der Spielraumplan angewandt werden, um das früher beschriebene Problem mit dem Halten der Ladung wettzumachen (Fig. 8B).

35 Wie schon vorher erwähnt, benutzen bekannte EEPROMs typischerweise  $V_{CC}$  zum Einspeisen von  $V_{CG}$  während des Programmier- oder Löschverifizierens. Um Spielräume vorsehen zu können, muß  $V_{CC}$  selbst angehoben oder reduziert werden. Durch diese Praxis werden in den Leseschaltkreisen ungenaue Ergebnisse erzeugt, da diese auch von  $V_{CC}$  angesteuert werden.

40 Bei der vorliegenden Erfindung ruft die Veränderlichkeit von  $V_{CG}$ , unabhängig von an die Lese- schaltkreise gelieferten Spannungen, genauere und zuverlässigere Ergebnisse hervor. Ferner ist die weite Spanne von  $V_{CG}$  beim Testen und bei der Diagnose der EEPROM nützlich. Sie macht es möglich, den vollen Bereich der Schwelle der programmierten Zelle leicht zu messen, indem

die Steigerung von  $V_{CG}$  fortgesetzt wird (bis zu dem durch den Verbindungszusammenbruch der Vorrichtung begrenzten Maximum).

5 Während die Ausführungsbeispiele dieser Erfindung, die hier beschrieben worden sind, die bevorzugten Verwirklichungen sind, ist dem Fachmann klar, daß auch Abwandlungen derselben möglich sind. Deshalb steht der Erfindung der Schutz innerhalb des vollen Umfangs der beige-fügten Ansprüche zu.

EP 0 539 358

Patentansprüche:

- 5        1. Nichtflüchtiger Speicher (130, 140, 150, 160), umfassend:  
ein Feld einer Vielzahl adressierbarer, elektrisch löscherbarer und programmierbarer Halbleiter-Speicherzellen der Art, die eine Source, eine Drain, ein Steuergate, ein schwimmendes Gate, welches in der Lage ist, einen bei der Benutzung des Speichers programmierten Ladungspiegel zu halten, und eine Löschelektrode aufweist, welche in der Lage ist, Ladung von dem schwimmenden Gate zu entfernen; ein System zum Lesen (220, 410, 420, 440, 513), ob der programmierte Zustand einer adressierten Zelle größer oder kleiner ist als ein vorbestimmter Schwellenwert; eine Referenzspeicherzelle (400, 529, 525); und eine auf die Referenzzelle reagierende Leseeinrichtung (410) zum Vergleich des Ladungspiegels einer adressierten Zelle mit demjenigen der Referenzzelle; gekennzeichnet durch  
eine auf das Löschen und Neuprogrammieren des Speicherzellenfeldes reagierende Einrichtung (210, 411) zum Löschen und Neuprogrammieren der Referenzspeicherzelle mit einer Ladung, die im wesentlichen gleich oder proportional dem Schwellenwert ist.
- 20        2. Nichtflüchtiger Speicher nach Anspruch 1, zusätzlich gekennzeichnet durch wenigstens eine Master-Referenzzelle, die mit einer Ladung löscherbar und programmierbar ist, welche im wesentlichen gleich oder proportional dem Schwellenwert ist, wobei die Einrichtung zum Neuprogrammieren der Referenzzelle eine Einrichtung zum Programmieren der Referenzzelle von der Master-Referenzzelle einschließt.
- 25        3. Nichtflüchtiger Speicher nach Anspruch 2, dadurch gekennzeichnet, daß die Leseeinrichtung eine Einrichtung zum direkten Vergleich des Ladungspiegels einer adressierten Zelle mit demjenigen der Referenzzelle einschließt.
- 30        4. Nichtflüchtiger Speicher nach Anspruch 2, dadurch gekennzeichnet, daß die Leseeinrichtung eine Einrichtung zur Einstellung des vorbestimmten Schwellenwerts der Master-Referenzzelle derart einschließt, daß dieser im wesentlichen dem der Referenzzelle angepaßt ist, und eine Einrichtung zum Vergleich der adressierten Zelle mit dem eingestellten Schwellenwert der Master-Referenzzelle.
- 35        5. Nichtflüchtiger Speicher (130, 140, 150, 160) umfassend:  
ein EEPROM-Feld, das eine Vielzahl von Gruppen (501, 503, 505) einzeln adressierbarer EEPROM-Speicherzellen enthält, wobei jede Zellengruppe zusammen als eine Einheit löscherbar ist,  
40        wenigstens einen vorbestimmten Schwellenpegel zur Abgrenzung zwischen zwei Leitungszuständen der Zellen in dem EEPROM-Feld,  
wenigstens eine Gruppenreferenzzelle (525), die als Teil jeder Speicherzellengruppe (503) zur Speicherung eines dem vorbestimmten Schwellenpegel entsprechenden Gruppenschwellenpegels vorgesehen ist,

eine Einrichtung zum Lesen (410, 440, 513) einer adressierten Zelle der Zellengruppe durch Vergleich mit dem Gruppenschwellenpegel,

5 eine auf das Löschen und Neuprogrammieren der Zellengruppe reagierende Einrichtung (210, 411) zum Löschen und Neuprogrammieren der Gruppenreferenzzelle auf einen dem vorbestimmten Schwellenpegel entsprechenden Gruppenschwellenpegel.

6. Nichtflüchtiger Speicher nach Anspruch 5, bei dem die Leseeinrichtung eine Einrichtung zum direkten Vergleich einer adressierten Zelle (523) mit den Gruppenreferenzzellen (525) der gegebenen Gruppe enthält.

10 7. Nichtflüchtiger Speicher nach Anspruch 5, enthaltend wenigstens eine Master-Referenzzelle zur Speicherung eines Master-Schwellenpegels (531), wobei die Leseeinrichtung eine Einrichtung zur Einstellung des Master-Schwellenpegels derart, daß er dem Gruppenschwellenpegel entspricht, und eine Einrichtung zum Vergleich des Schwellenpegels einer adressierten 15 Zelle (523) mit dem so eingestellten Master-Schwellenpegel enthält.

20 8. Nichtflüchtiger Speicher nach den Ansprüchen 1, 5, 6 und 7, ferner umfassend ein System zum Programmieren der Speicherzellen, dadurch gekennzeichnet, daß die Leseeinrichtung auch Teil des Systems zum Programmieren der Speicherzellen ist, bei dem jede adressierte Speicherzelle zu einem gewünschten Zustand programmiert wird, indem eine Folge des Programmierens mit einem kleinen Schritt und Lesens zum Verifizieren des programmierten Zustands wiederholt wird, bis der gewünschte Zustand erreicht ist.

25 9. Nichtflüchtiger Speicher nach den Ansprüchen 1, 5, 6 und 7, ferner umfassend ein System zum Löschen der Speicherzellen, dadurch gekennzeichnet, daß die Leseeinrichtung auch Teil des Systems zum Löschen ist, bei dem jede adressierte Zelle zu dem Löschkzustand gelöscht wird, indem eine Folge des Löschens in einem kleinen Schritt und Lesens zum Verifizieren des Löschkzustands wiederholt wird, bis der Löschkzustand erreicht ist.

30 10. Nichtflüchtiger Speicher nach Anspruch 2, dadurch gekennzeichnet, daß der in jeder genannten Master-Referenzzelle gespeicherte Ladungspegel von außerhalb des Speicher- systems elektrisch lösbar und programmierbar ist.

35 11. Nichtflüchtiger Speicher nach Anspruch 5, ferner gekennzeichnet durch einen Satz Master-Referenzzellen (507, 529), der den EEPROM-Feld zur Speicherung des Satzes vorbestimmter Schwellenpegel zugeordnet ist.

12. Speichersystem nach Anspruch 11, dadurch gekennzeichnet, daß die Master-Referenzzelle ein Element der Speicherzellen des Feldes ist.

40 13. Nichtflüchtiger Speicher nach den Ansprüchen 1 und 2-4 sowie 10, dadurch gekennzeichnet, daß die Referenzzelle ein Element der Speicherzellen des Feldes ist.

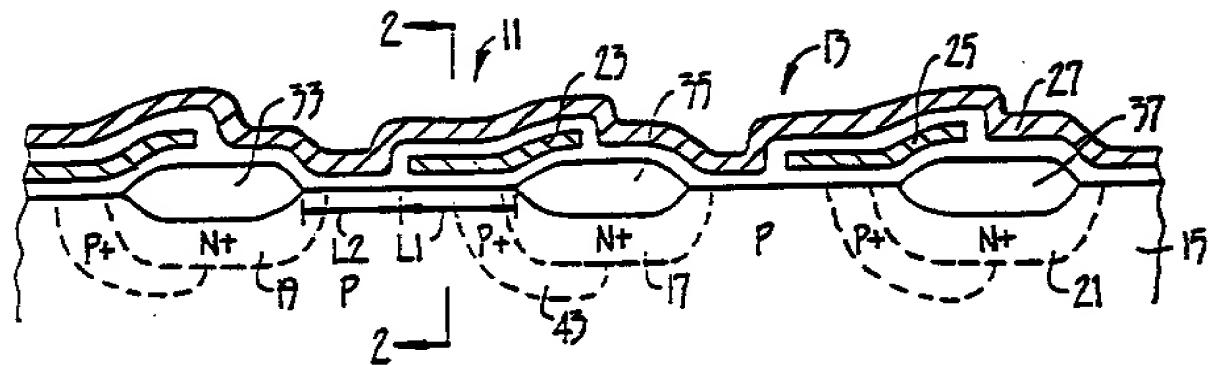


FIG. 1.

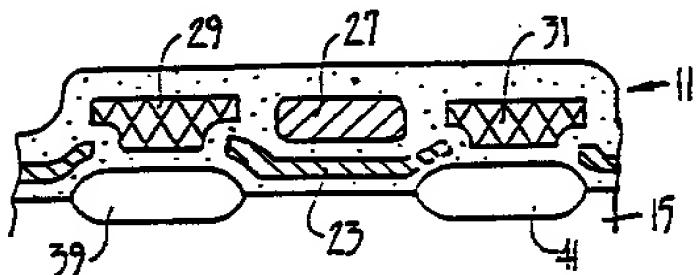


FIG. 2.

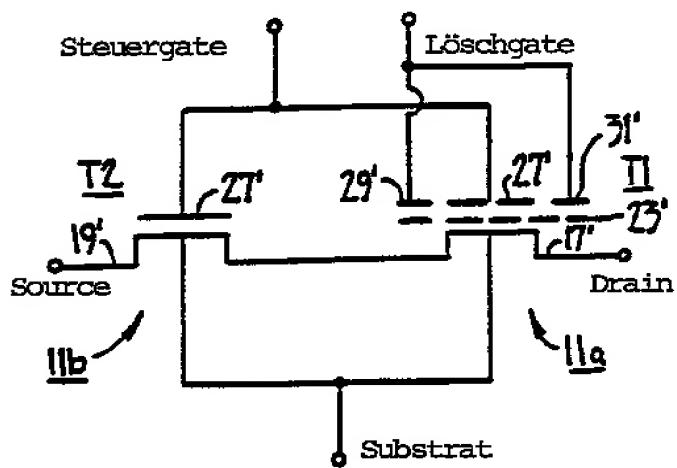
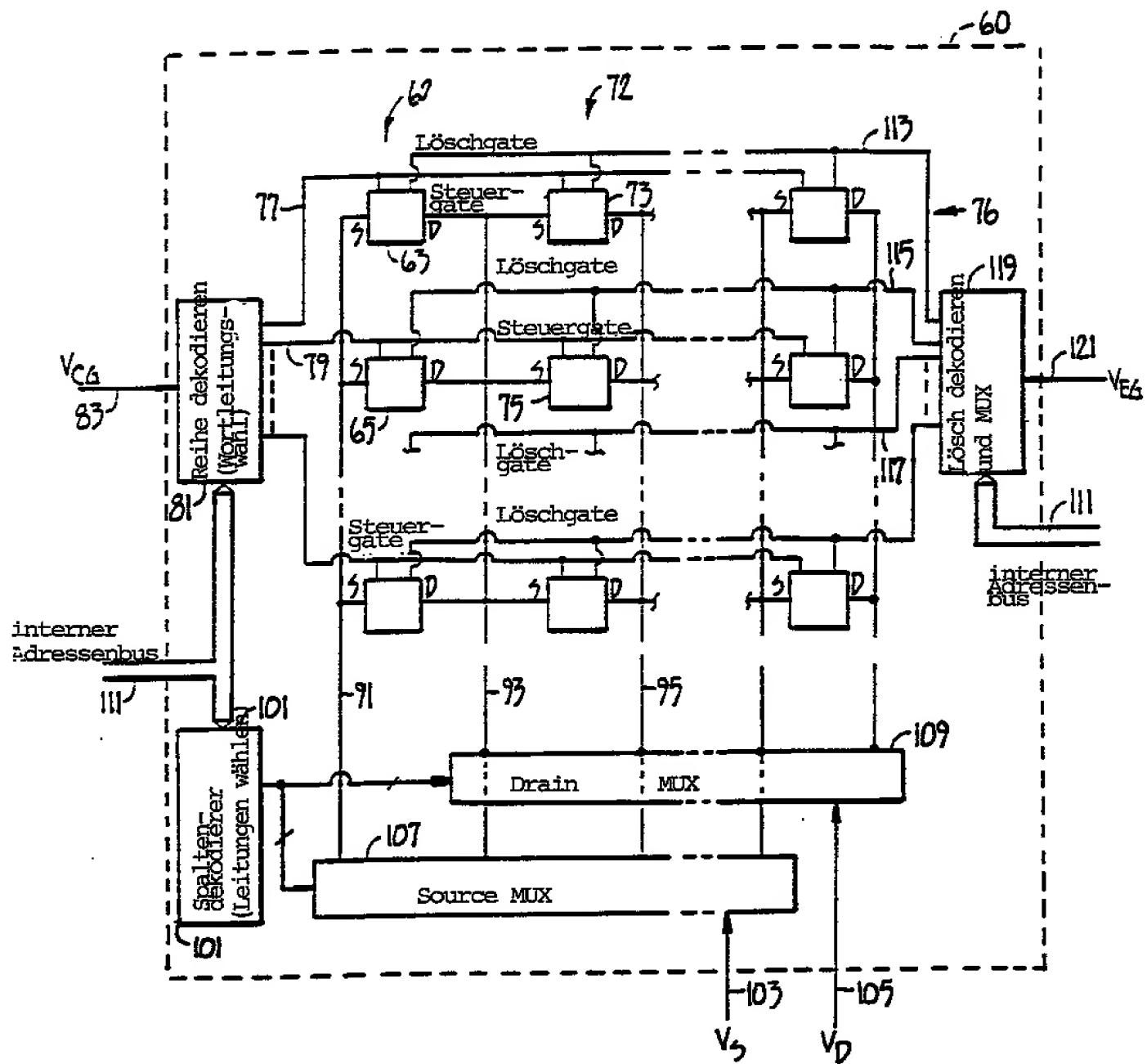


FIG. 3.



FIG\_4.

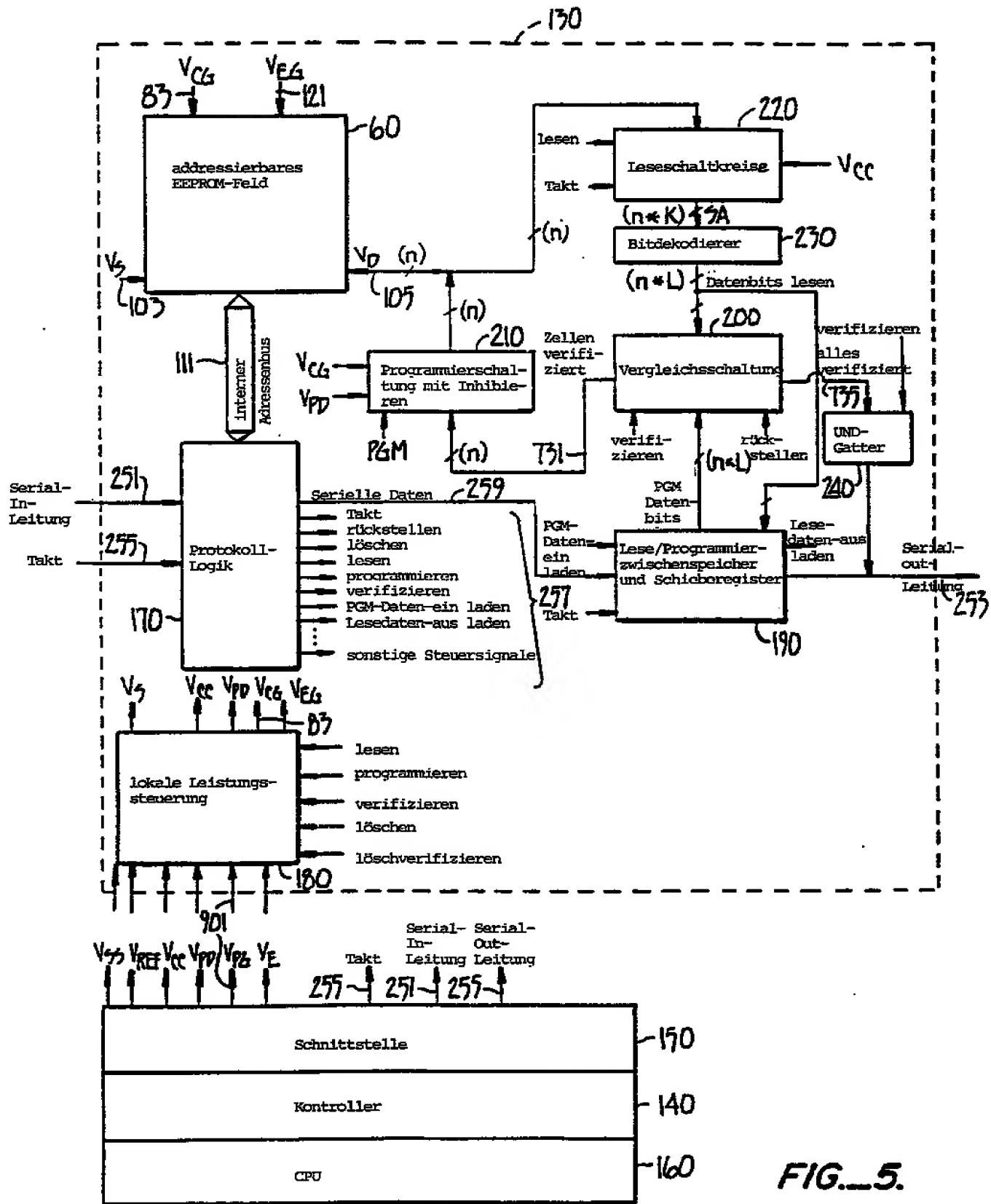


FIG.—5.

4/17

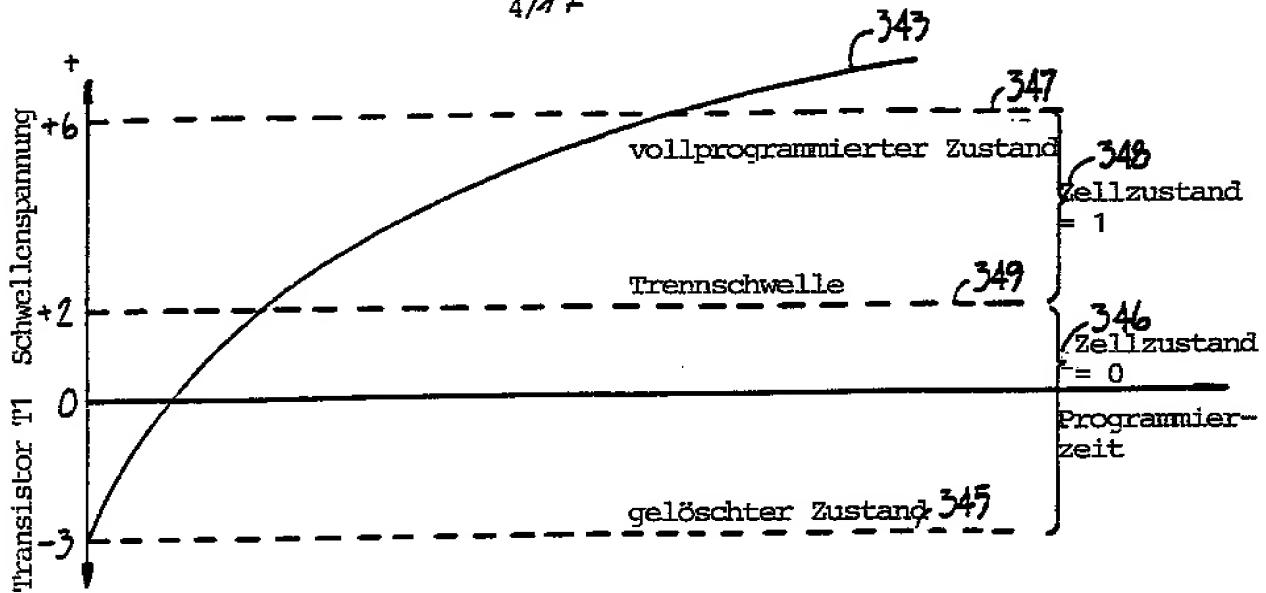


FIG. 6.

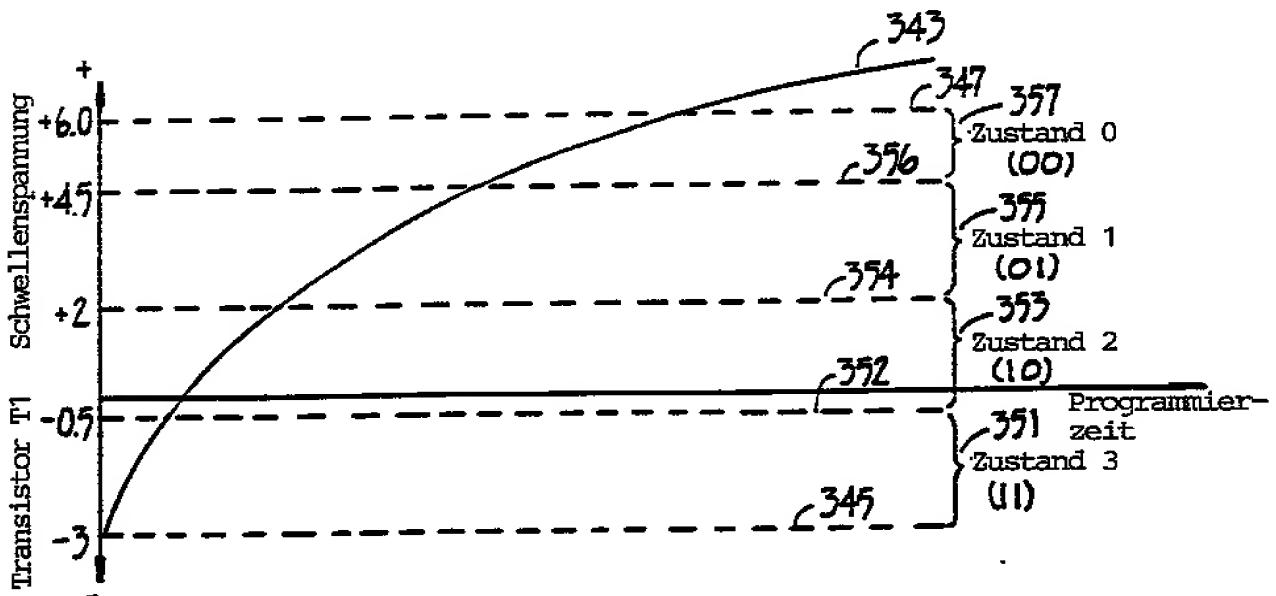
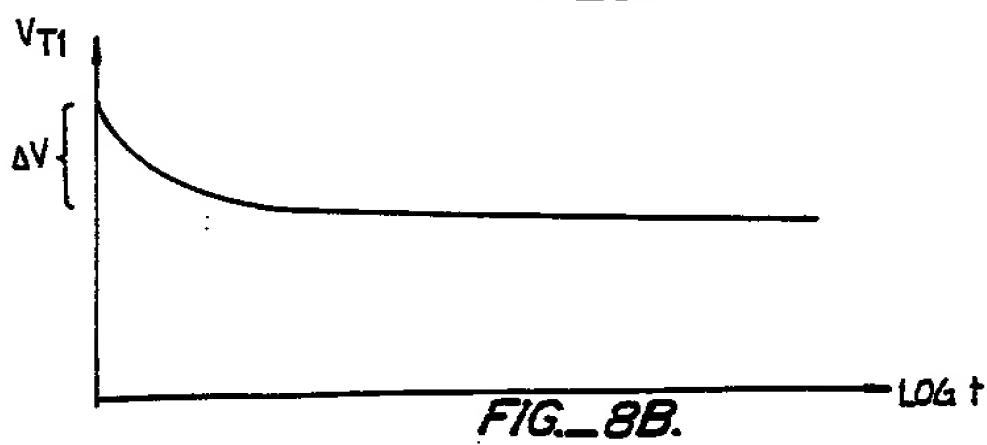
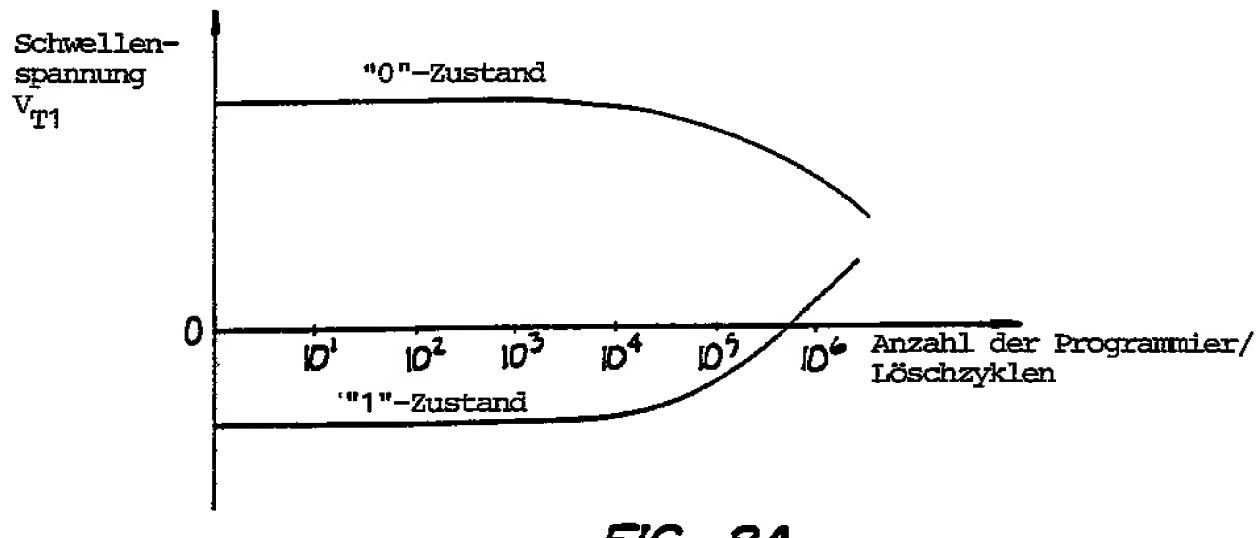
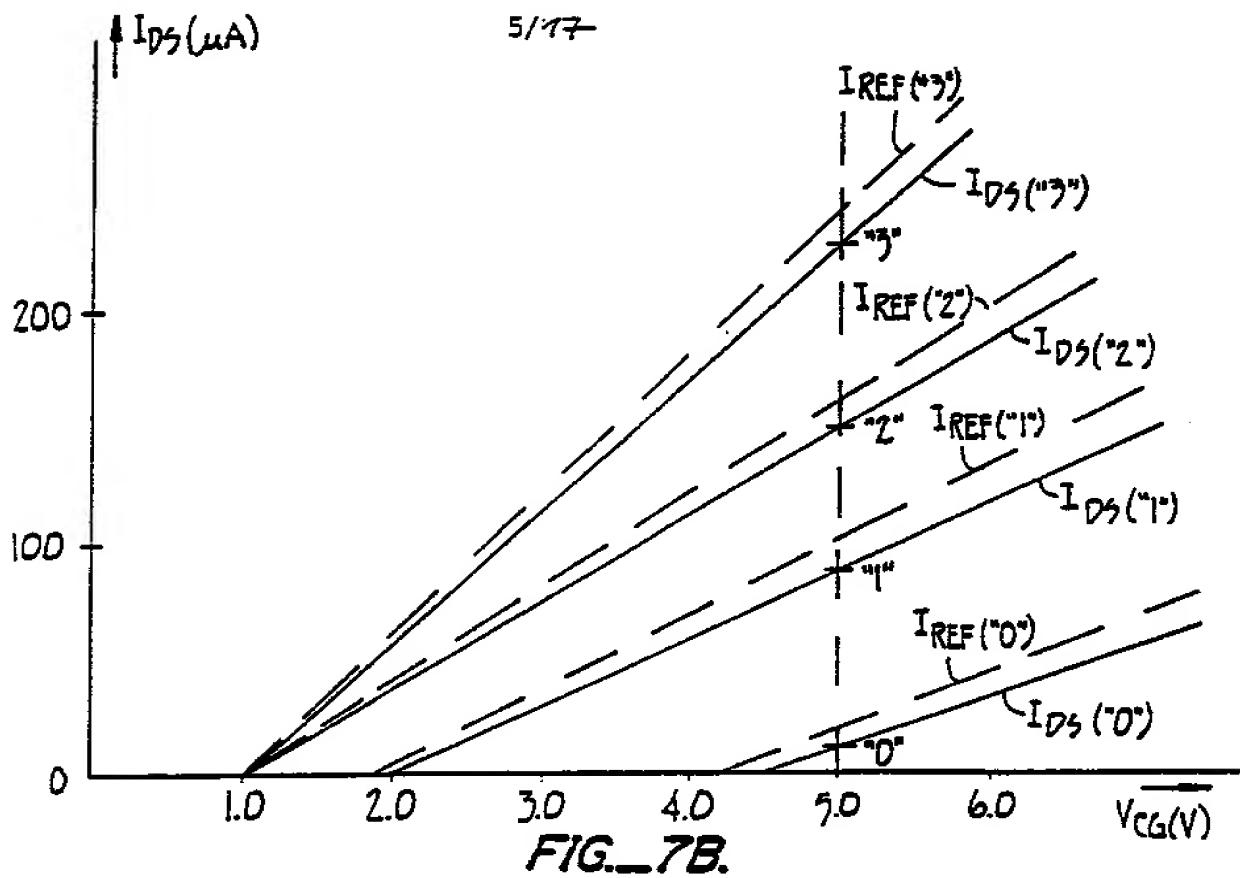


FIG. 7A.



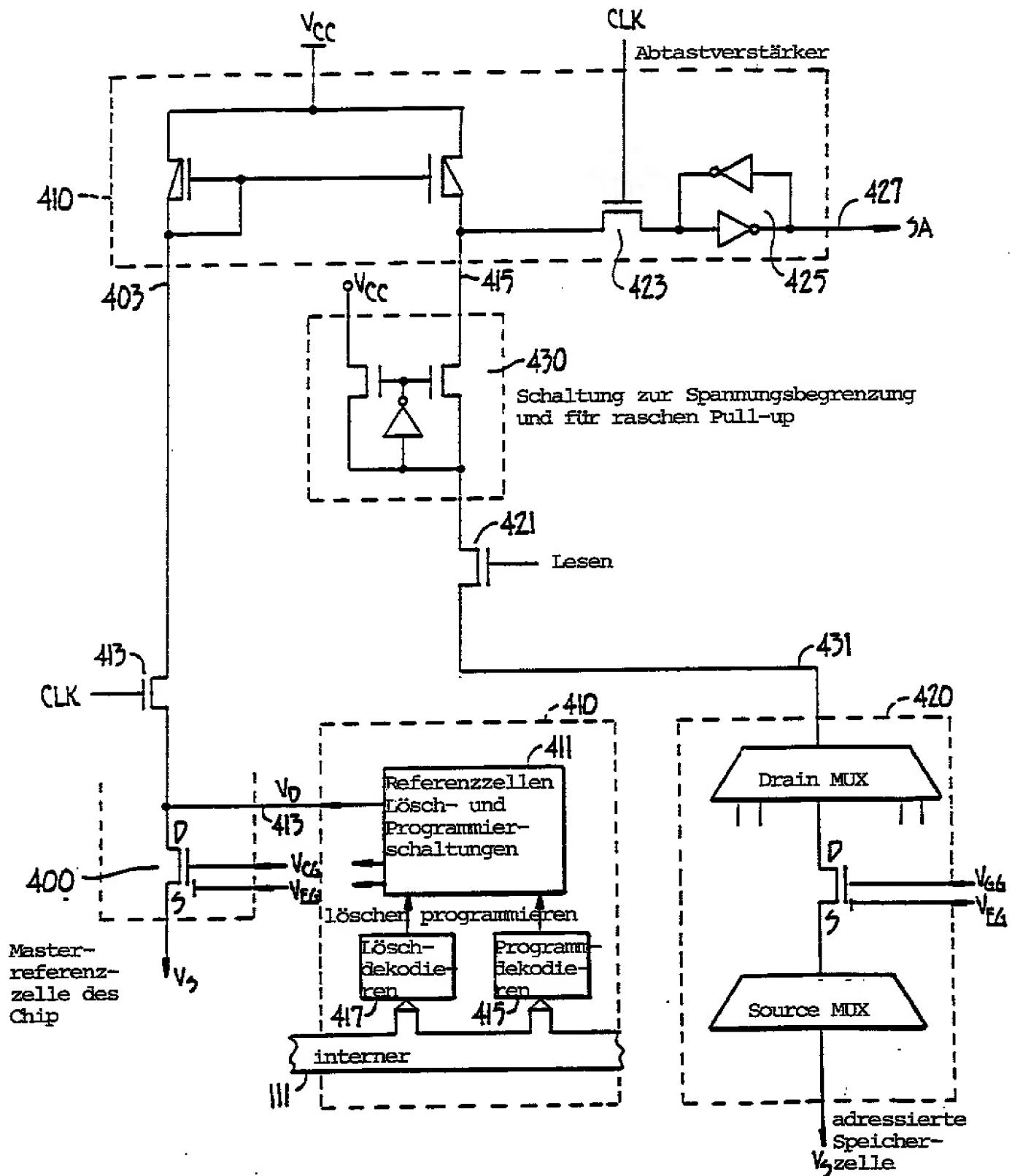


FIG. 9A.

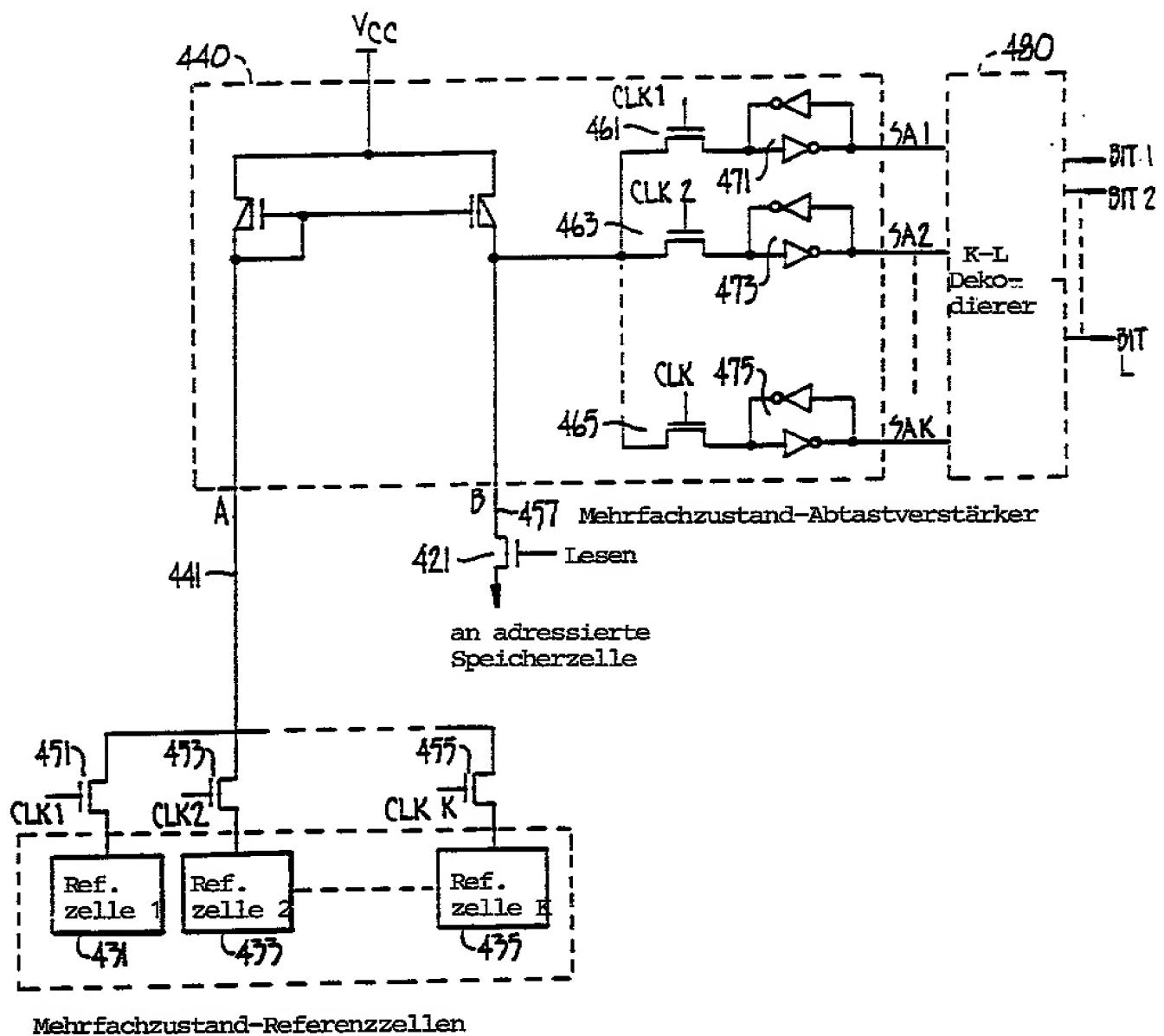


FIG. 9B.

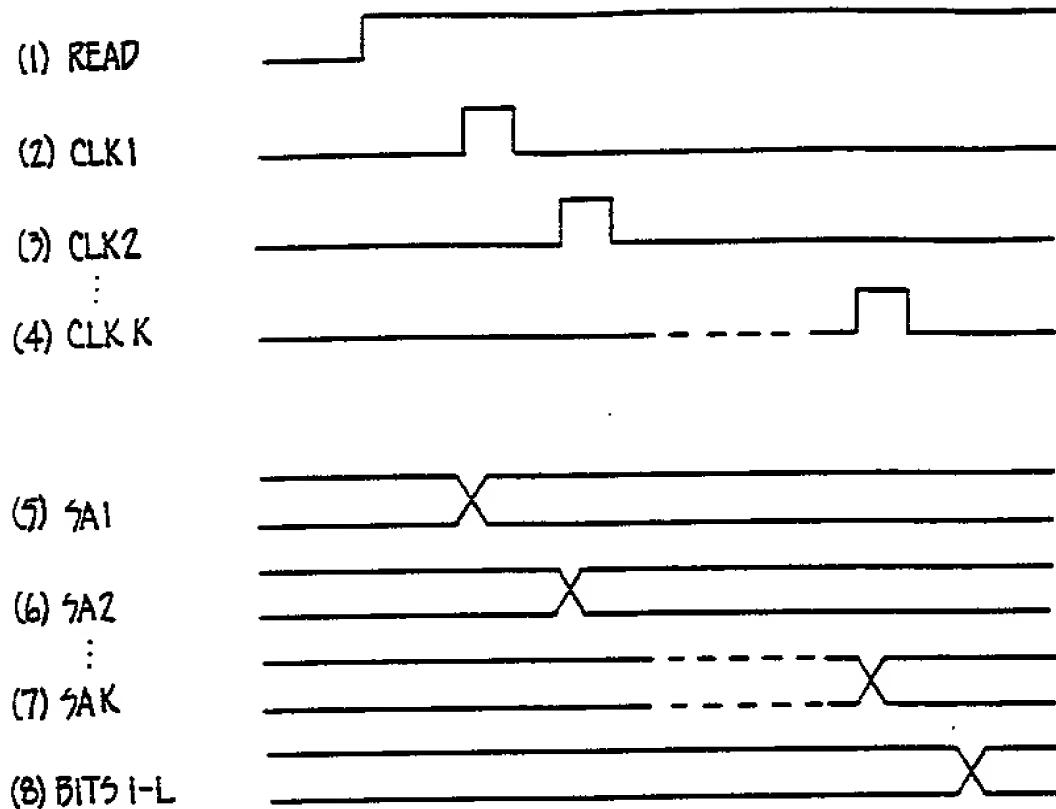


FIG. 9C.

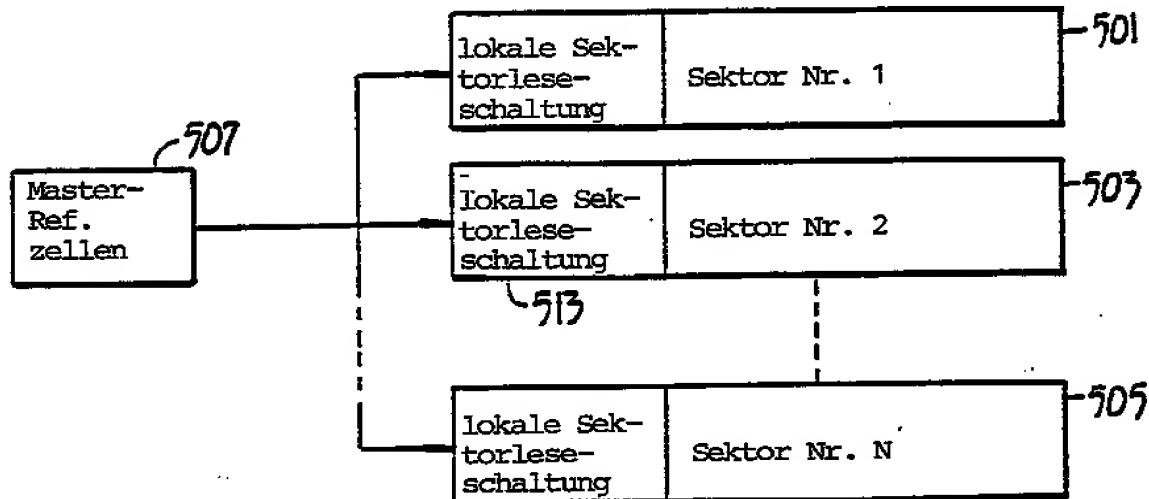
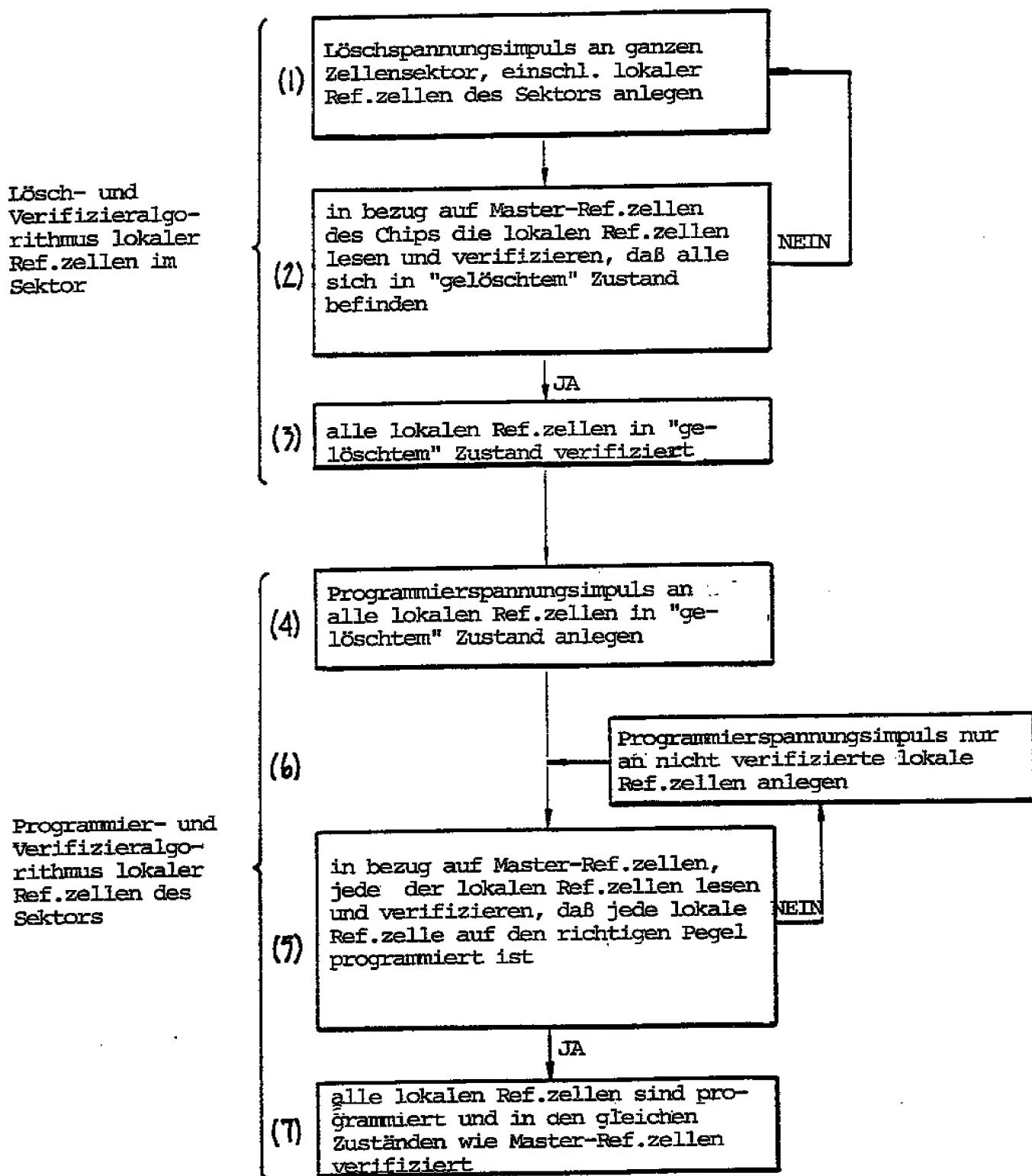


FIG. 10.



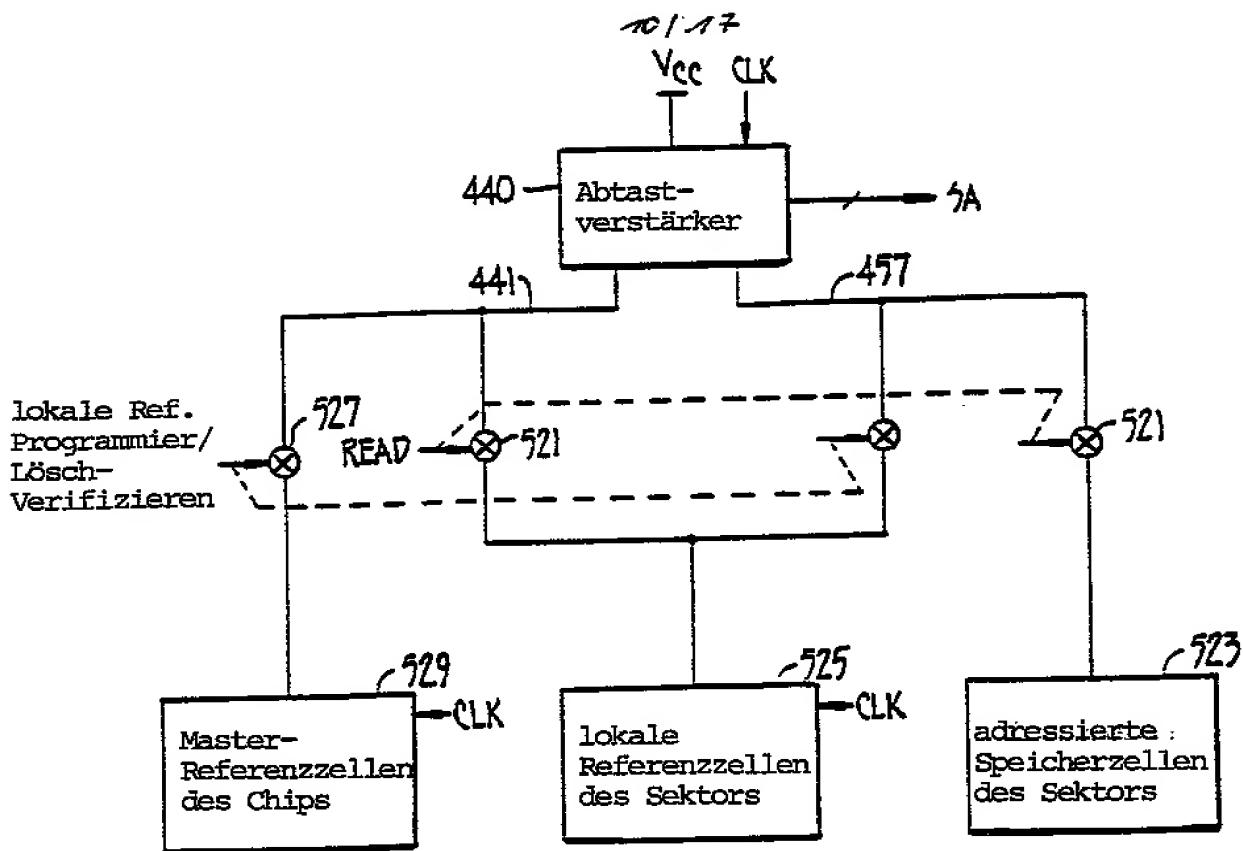


FIG. 12A.

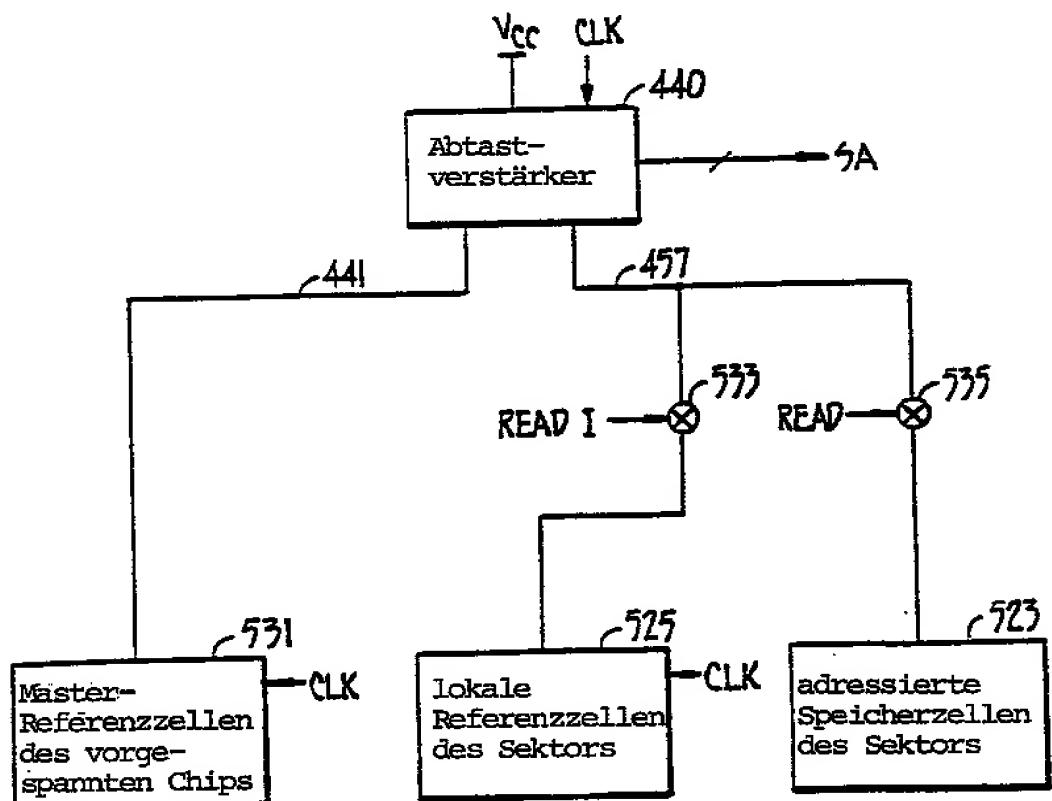


FIG. 13A.

11/17

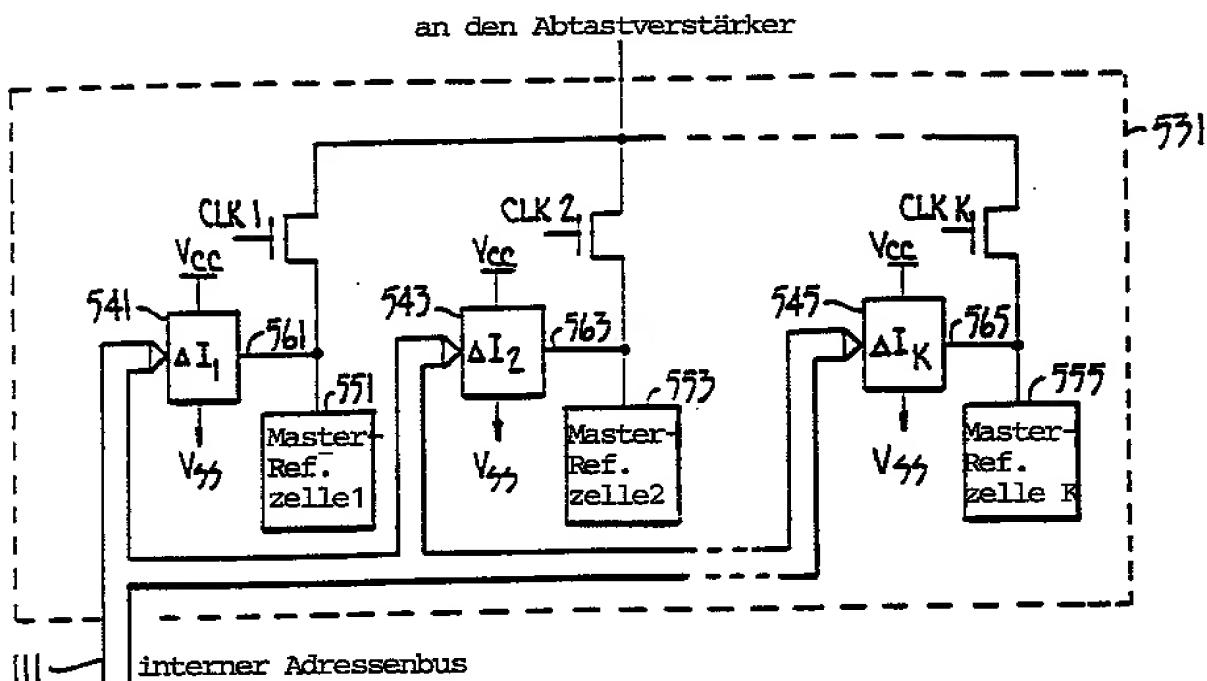


FIG. 13B.

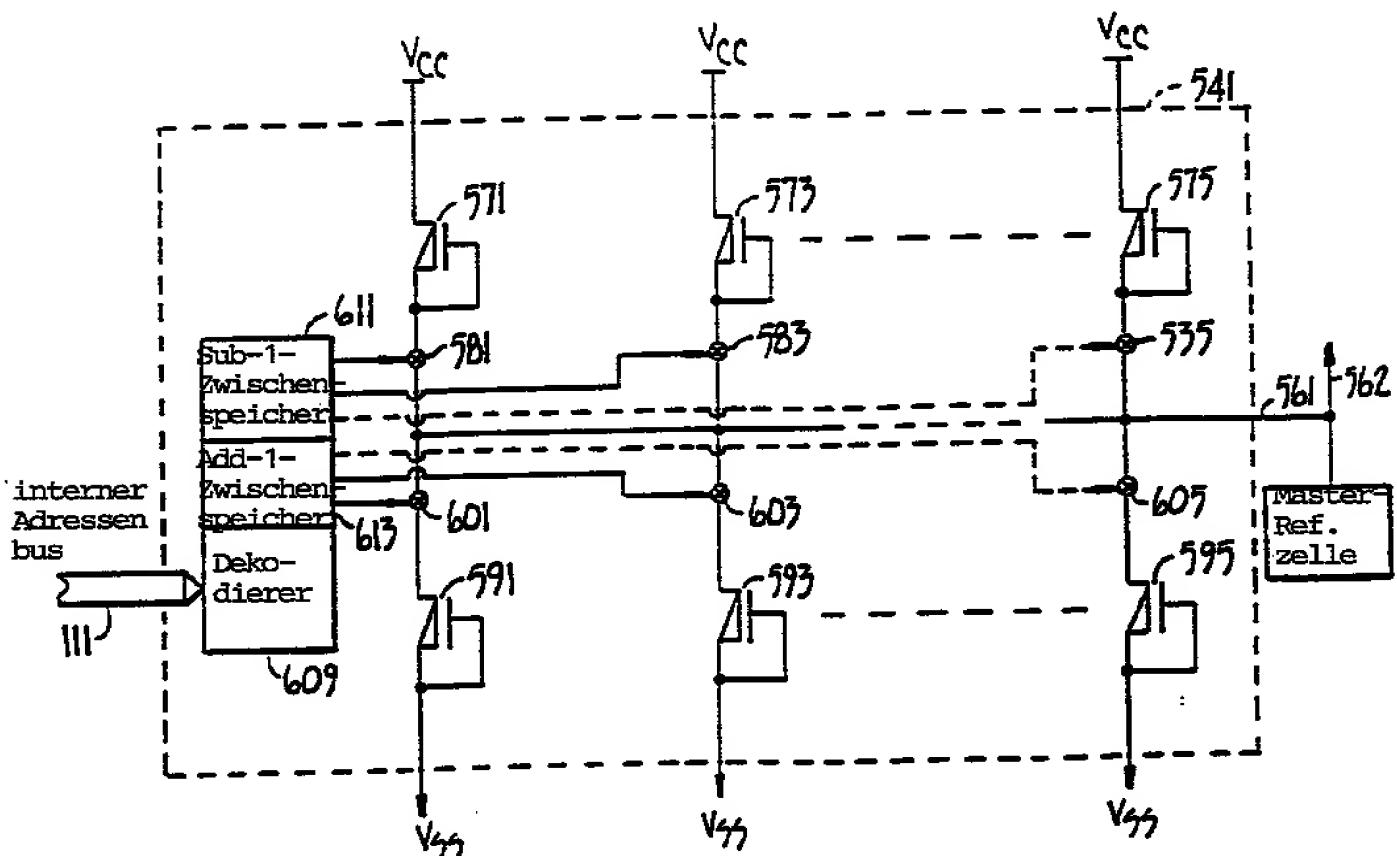


FIG. 13C.

12/17

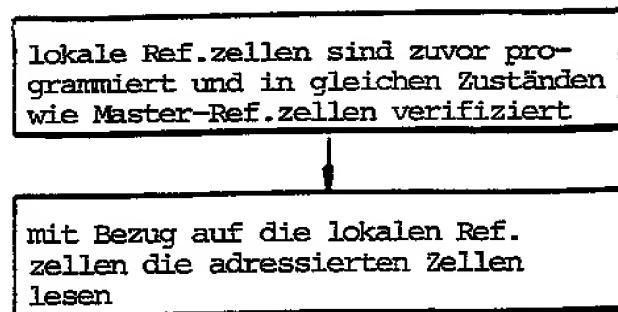


FIG.\_12B.

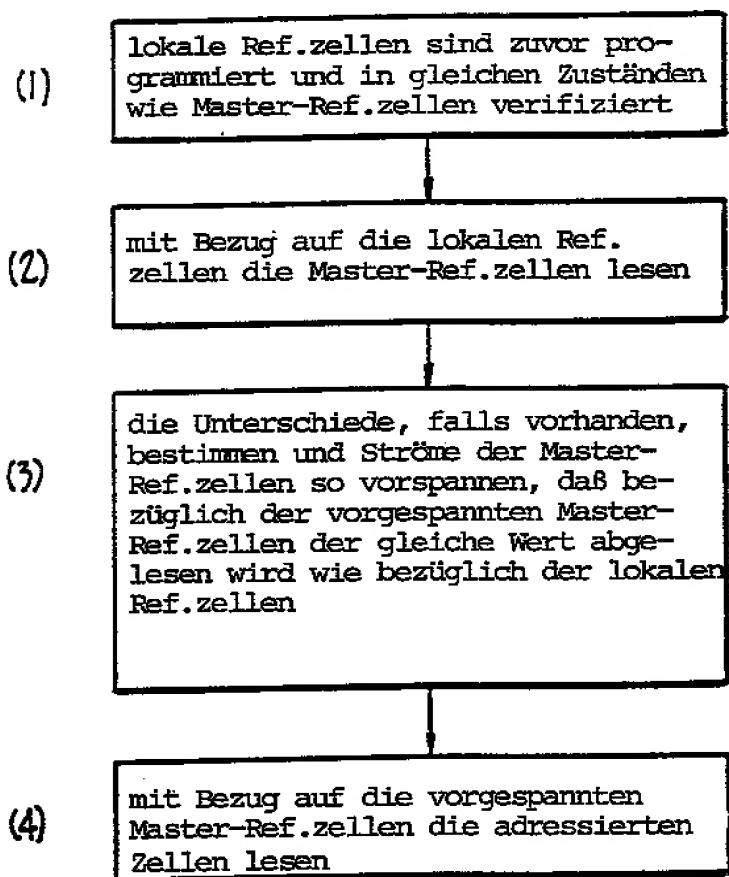
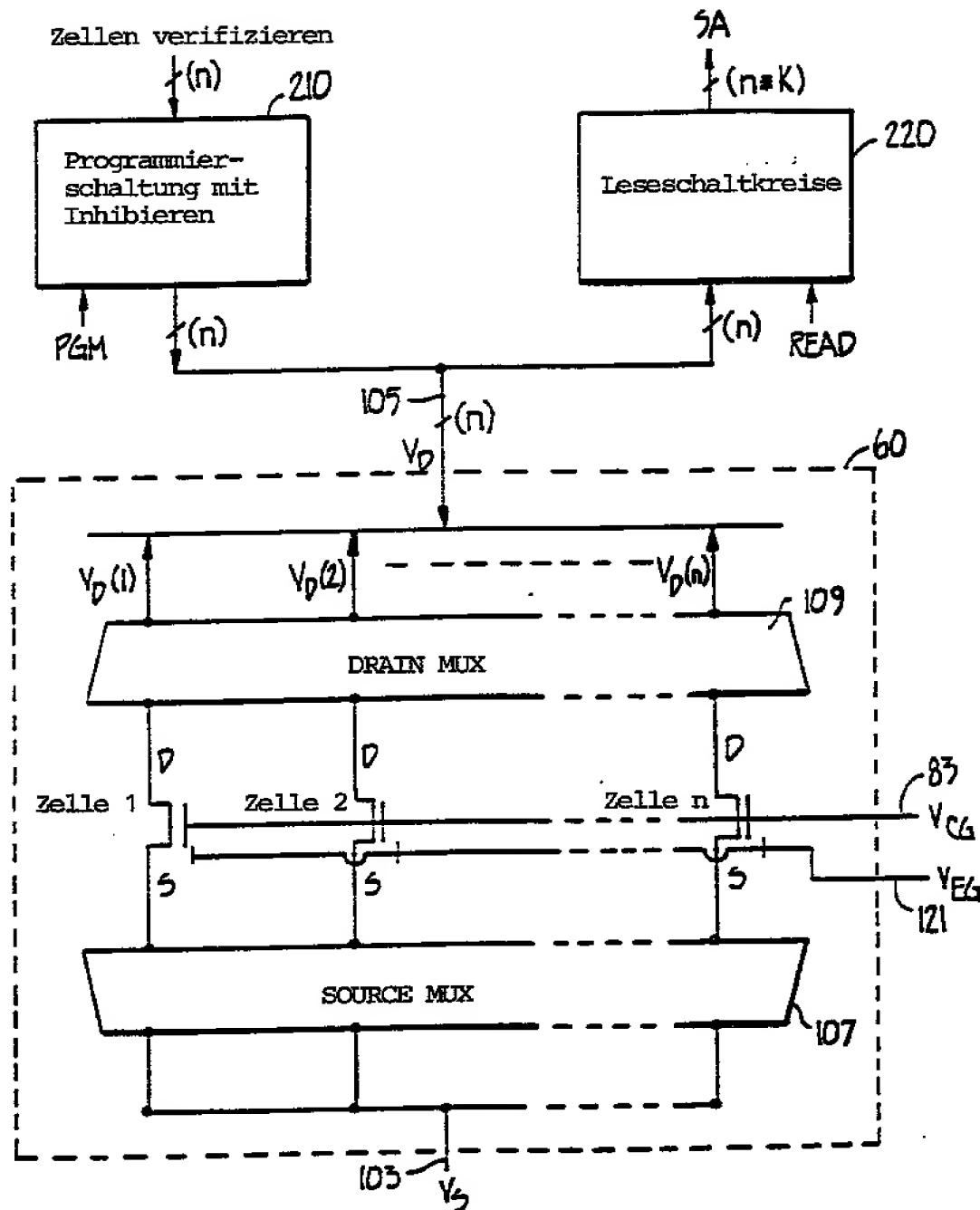


FIG.\_13D.



## Lese/Programmier-Datenwege für n parallele Zellen

FIG. 14.

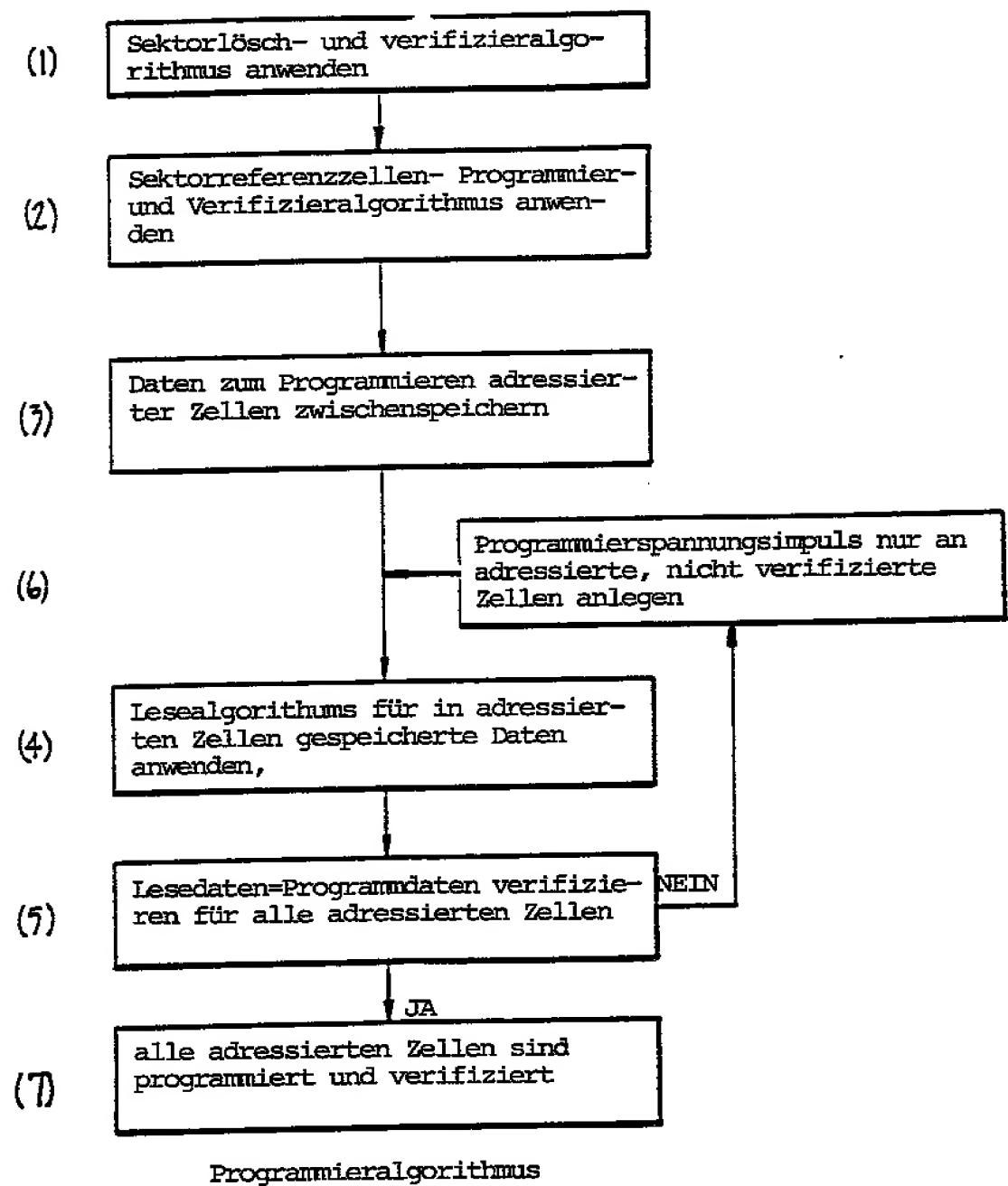


FIG. 15.

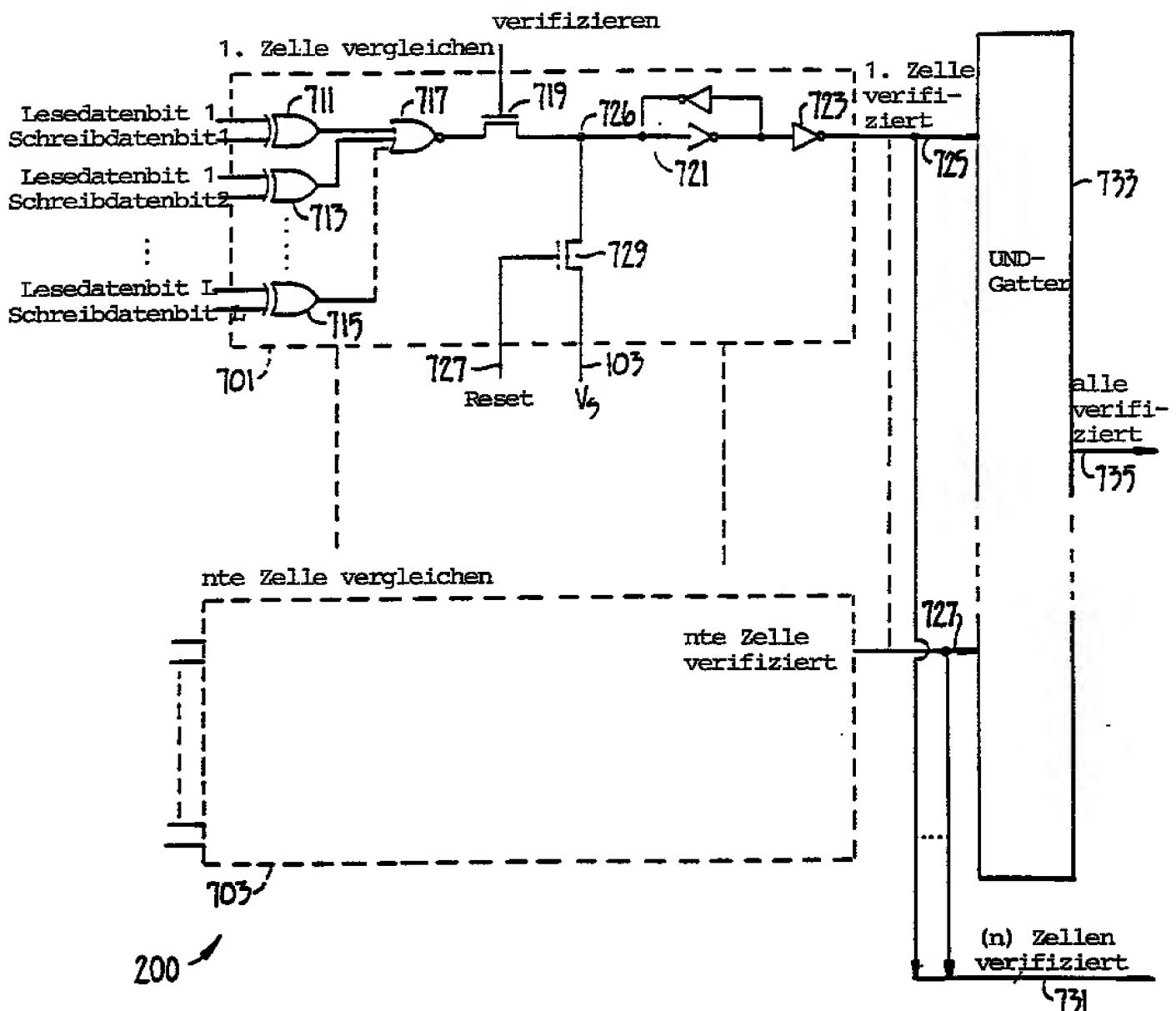


FIG.-16.

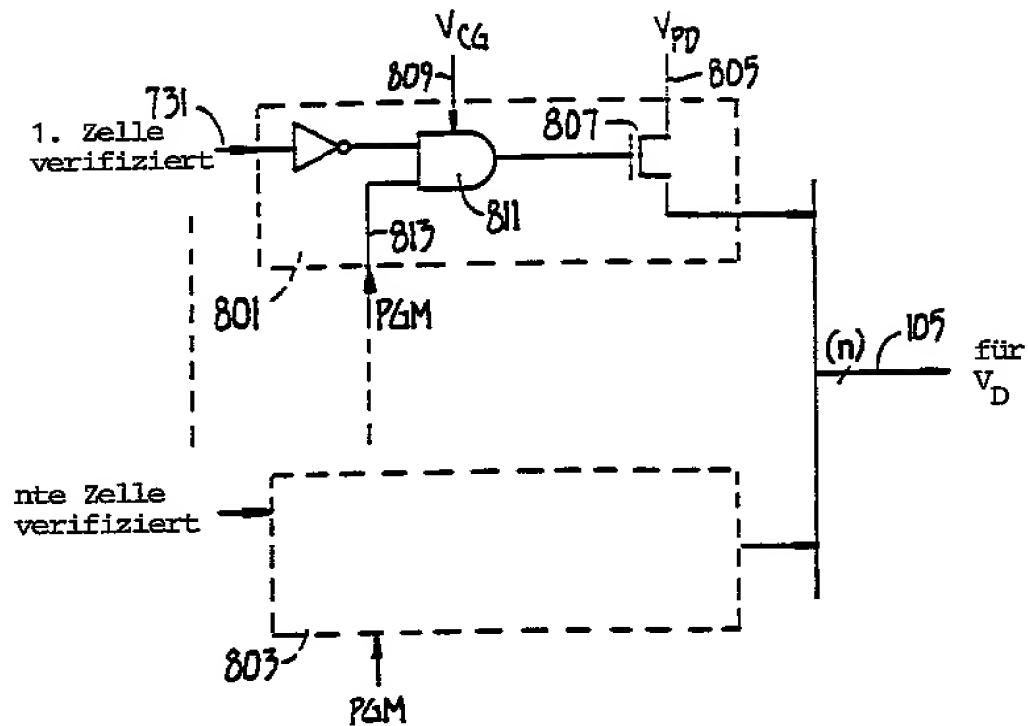


FIG. 17.

	Gewähltes Steuergate $V_{CG}$	DRAIN $V_D$	SOURCE $V_S$	Löschgate $V_{EG}$
lesen	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
programmie- ren	$V_{PG}$	$V_{PP}$	$V_{SS}$	$V_E$
programm- verifizieren	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
löschen	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$
löschveri- fizieren	$V_{PG}$	$V_{REF}$	$V_{SS}$	$V_E$

TABLE 1

(Typische Werte)	lesen	programmie- ren	programm- verifizie- ren	löschen	löschverifi- zieren
$V_{PG}$	$V_{CC}$	12V	$V_{CC} + 5V$	$V_{CC}$	$V_{CC} - 5V$
$V_{CC}$	5V	5V	5V	5V	5V
$V_{PD}$	$V_{SS}$	8V	8V	$V_{SS}$	$V_{SS}$
$V_E$	$V_{SS}$	$V_{SS}$	$V_{SS}$	20V	$V_{SS}$
nicht gewähltes Steuergate	$V_{SS}$	$V_{SS}$	$V_{SS}$	$V_{SS}$	$V_{SS}$
nicht gewählte Bitleitung	$V_{REF}$	$V_{REF}$	$V_{REF}$	$V_{REF}$	$V_{REF}$

$$V_{SS} = 0V, \quad V_{REF} = 1.5V, \quad 5V = 0.5V - 1V$$

TABLE 2